

Proper priorities needed for the NPT

Everybody's interest is that North Korea should not be aggressively pilloried for its less than full cooperation with the recent inspection of its nuclear plants; the survival of the Non-Proliferation Treaty is more important.

IN just about a year, governments that are members of the Nuclear Non-Proliferation Treaty (NPT) should know the agenda of the conference later next summer to mark the twenty-fifth anniversary of the ratification of the NPT. It will be a crucial meeting. The text requires a positive decision by the existing members that the treaty should remain in force beyond next year. It is in everybody's interest that the decision should be to keep the treaty going.

As an instrument of international peace of mind, the NPT has been remarkably successful. The number of overt nuclear powers is no greater now than 25 years ago. And although there may be a number of covert nuclear powers (Israel, India and possibly Pakistan), others (Argentina and Brazil, for example) have disappeared from the lists. But more than that, the NPT has been one of several spurs towards the measures of strategic arms control the United States and the former Soviet Union have agreed; the latest development there is a remarkable agreement by which the two sides will actually count the plutonium triggers removed from each other's dismantled hydrogen-bomb warheads; that may be a basis for removing from a future version of the treaty the asymmetry by which nuclear powers are free to inspect each other while enforcing inspection on non-nuclear powers. It is to be hoped that next year's agenda will also include proposals on international custody of spent fuel and the like, but that is something else.

The danger is that the atmosphere for an agreement next year may be ruined by a row over North Korea, consistently (and almost mischievously) over the past two years a black sheep among NPT members, the example of Iraq notwithstanding. First, North Korea gave notice of its intention to withdraw from the NPT (of which it is a signatory), then it withdrew its withdrawal under diplomatic pressure, mostly from the United States, but appears to have forbidden a team of inspectors sent by the International Atomic Energy Agency to collect samples of plutonium from a radioactive glovebox at one nuclear site. The isotope composition of plutonium, which is easily determined, is well-known to be a test of whether the material has been bred to make bombs. The inference from North Korea's refusal is that its government has something to hide.

So there have been demands, notably in the United States, for economic sanctions against North Korea. The US Secretary of State, Mr Warren Christopher, has even given his view that China would probably not veto such a proposal, even if it failed to support it. Sadly, Christopher's frustrating

experience in Beijing earlier last week is hardly such as to commend his opinion of what China will or will not do.

But in any case, given that next year's NPT conference must have it as a central objective to cement Chinese support for the NPT regime, even to risk a Chinese veto in the UN Security Council would dangerously divide the world's nuclear policemen (of which China must be one) on the eve of a conference where it is essential that they act in unison. That is a more important objective than North Korea's instant compliance with the inspection procedures. After all, despite talk of North Korean development in ballistic missiles as well as bomb technology, nobody seriously pretends that North Korea is now in a position to be a military threat to its neighbours — or that China would allow such uppishness. In due course, North Korea must be compelled to comply with the inspection regime it has invited, but too big a stick at this point could do more harm than good. □

Efficiency in research

The Royal Society has fired a warning shot at the planned review of British research establishments.

ONCE upon a time, in the early decades of this century, the British government made several important innovations in the conduct of research, partly out of the recognition that it was technically ill-prepared for the First World War. A quarter of a century later, when the Second World War began, the tradition that it is an important function of government to carry out research was well-established. Many of the laboratories thus created, the radar establishment at Malvern (now part of the Defence Research Agency) performed superbly and won illustrious reputations. By the 1950s, there was a tendency to create a new research establishment for each new problem or opportunity that arose. Since 1970, the tendency has been reversed; laboratories have been amalgamated or even closed. And now the government has embarked on what is called an "efficiency scrutiny" of 53 public laboratories supported by a variety of government departments (from agriculture to industry) but excluding defence.

The good news is that the newly created Office of Science and Technology (OST) is closely involved in the exercise. Whatever the detailed conclusions of the scrutiny, its general effect may be to give that office a continuing role in the



supervision of government departments' commitment to research. That would be a decisive break from the ethos of the Rothschild Report, which decreed in 1971 that all departments should equip themselves with a chief scientist capable of being an effective "customer" for research, but it is a necessary break: departments have since been pusillanimous in their regard for research, all too ready to count it as a dispensable cost when budgets are tight. If Britain ends up with a Civil Research Agency responsible to OST, that will be entirely welcome.

The worries are different. The "scrutiny" is being conducted by civil servant members of the government's "Efficiency Unit", who are not chosen for their feeling for research, has been done in a hurry (90 days) and is likely to have been driven by the fashionable view that privatization in some form (outright sale or management by private contractors) is inherently virtuous. The Royal Society is right to fear that the outcome may pay scant regard to the principle that much research is not a commodity but an activity that is valueless without continuity. Without its Central Veterinary Laboratory (one of the 53 being scrutinized), the government would be in an even greater pickle in knowing how to deal with the current outbreak of bovine spongiform encephalopathy (BSE), for example. And is it possible to run an effective laboratory for the management of safety standards for vaccines without a continuing fund of expertise in the field?

That is one cause for alarm. The Royal Society also fears that when the scrutiny report is published, the government will already have made up its mind. The assurance from Mr William Waldegrave, the science minister, that consultation will be genuine is therefore welcome, brief though the period allowed (another 90 days) may be. The more enduring cause for concern is that the outcome of the scrutiny, if it should allow government departments to indulge their inclination to wash their hands of responsibility for research, will leave the British Civil Service even less well-equipped than at present to fight its government's corner in an increasingly technical world, in Brussels and such places. OST will have its work cut out to shoulder that whole burden, but there may be no choice. □

Britain's block votes

The British government is once again in the European doghouse, this time for good reason.

How should the interests of minorities be safeguarded in circumstances where decisions are normally made by counting votes? The question often arises, and is answered differently in different circumstances. In many countries, Russia for example, a popular vote on an amendment to the constitution requires a two-thirds majority to succeed. In many places, juries in court trials are required to give verdicts that are unanimous, but elsewhere a specified majority (say 10 out of 12) will suffice. On the United Nations Security Council, any one of the five permanent member govern-

ments can unilaterally veto a decision it does not like.

It used to be like that on what is called the European Council, at which representatives of member governments, usually ministers, adopt (or otherwise) European legislation. But by the common consent of the 12 present members, the rules were changed in 1987 to weight the votes of different member states crudely by their population (the four large members have 10 votes each, smaller member states have 4, 3 or even 2, making a total of 76). At the same time, it was agreed that a minority seeking to block the adoption of proposed legislation must assemble 23 votes, or roughly 30 per cent, to succeed. None of this touches the vast range of European policy-making concerned with foreign policy, security and defence, where the assertion of an "overriding national interest" will suffice to give a single member state a veto, as of old.

So what happens when new members join the European Union (EU)? (The good news that Austria, Finland and Sweden now wish to join has since been followed by an agreement with Norway.) They, too, must evidently have a weighted vote. The negotiations so far (to which Britain has been privy) would augment the total number of votes to be cast at a meeting of the European Council to 90 and the required blocking minority vote to 27, or 30 per cent of the new total. The British government, which has creditably been consistent in its backing for the enlargement of the EU, has now objected to the arithmetic, on the grounds that any increase of the number of votes required to block a piece of European legislation will further diminish the autonomy of member states.

Whether or not a compromise will have been reached at this week's meeting of foreign ministers in Brussels, the fuss the British are making is disingenuous humbug. (Spain also protests, but from fear that the outcome of the accession of the northern states will be a decrease of the funds available for Mediterranean agriculture, and in any case has done so less obdurately.) The British objection to the proposed voting regime springs not from an objective appraisal of what would be just, but from domestic political considerations: the government has only just emerged from a bruising battle with its own supporters over the ratification of the Maastricht Treaty, and wishes to be seen to be acting "tough" on Europe in advance of the elections to the European Parliament to be held throughout Europe in June. The circumstances have been widely reported, are generally understood and are discreditable.

In none of its calculations does the British government appear to have grasped that its wish to be "at the heart of Europe" (as the prime minister put it some months ago) is unlikely to be gratified by a row over the precise circumstances in which Britain (in concert with others) can vote down proposed legislation. It is especially disastrous that this should happen precisely when the British government should have been seeking to mend the damage done by the tawdry rhetoric of its internal squabble over Europe last year. If the outcome is that the government party's performance in the June elections turns out to be even worse than now predicted, that will be small comfort. □

US research said to have provided basis of Britain's hydrogen bomb

Washington. A new history of British, French and Chinese nuclear weapons suggests that the British nuclear programme depended considerably more heavily on US research, development and testing than previously acknowledged.

The authors of the 500-page survey, published in Washington today (24 March) by the Natural Resources Defense Council (NRDC), say that Britain's first two attempts to detonate a hydrogen bomb in May and June 1957 were unsuccessful, despite claims at the time in the British press and by prime minister Harold Macmillan.

The survey also says that Britain's own H-bomb design was discreetly abandoned after 1958, when the United States and United Kingdom resumed nuclear weapons collaboration. It claims that nuclear weapons subsequently deployed by Britain were "nearly direct copies" of US designs.

But some British historians dispute this theory, saying that it may be based on incorrect assumptions about how much testing is needed to perfect a nuclear weapon. They say that available knowledge of the earliest British nuclear weapons demonstrates a willingness to deploy weapons after only a

handful of tests, without the extensive testing programmes favoured by the United States and France.

The NRDC survey suggests that the WE 177 bomb, the Royal Air Force's main nuclear weapon since it entered service in 1966, is basically a UK-manufactured variant of a US design.

Survey co-author Stan Norris says there are several current theories about the origins

that the WE 177 "had 80–90 per cent commonality with the American bomb".

But John Simpson of the Mountbatten Centre for International Studies at Southampton University, a leading historian of the British atom bomb programme, disagrees. He still believes the WE 177 was a British design based mainly on tests conducted before a three-year nuclear test moratorium began in 1958. Simpson says the British were thinking about the WE 177 as early as 1957, and that they did not feel they needed the kind of test programmes that the United States conventionally used. "There was a view that where the Americans took seven or eight tests, the British would take one," he says.

However both Simpson and Norris acknowledge that they remain ignorant of what really happened. The intense secrecy that has surrounded the British H-bomb programme has begun to lift slightly; earlier this year, for example, the Public Records Office at Kew released data about testing under the 'thirty-year' rule on classified documents.

Key files on Britain's own programme up to 1958 were closed in 1964, and have therefore only just been released. But data on the British programme after 1958 is unlikely to follow, since neither Britain nor the United States is permitted to release data about each other's role in the collaboration.

The authors of the NRDC survey claim that it is the most comprehensive history and description of the UK weapons programme yet published — an opinion endorsed on the report's sleeve by former US defence secretary Robert McNamara.

Norris claims that Britain obtained its nuclear arsenal "on the cheap" after resuming weapons collaboration with the United States in 1958. Noting the secret problems that Britain had with its own hydrogen bomb tests in 1957, the survey claims that the British design was effectively shelved the following year.

He says that their explosive yields of 200–300 kT during tests in May and June 1957 indicate that the second, thermonuclear stage of the devices did not work. Only later, in November, did Britain succeed in detonating a genuinely thermonuclear device in which a large proportion of the energy released came from the fusion of hydrogen.

British, French and Chinese Nuclear Weapons is the fifth volume of NRDC's Nuclear Weapons Databook. The previous volumes revealed details of the larger nuclear weapons industries of the United States and the Soviet Union. **Colin Macilwain**



Britain's WE 177 bomb: based on US design?

of the WE 177. He argues that the paucity of UK tests — only four were held in the crucial period 1962–1965 — undermines the claims that the bomb was an original British design.

Instead, he suggests that successive versions of the WE 177 were based on the American B57 and B61 weapons respectively. In quantitative terms he estimates

Hammersmith celebrates a reprieve

London. The British government has backed away from a decision to merge the Hammersmith Hospital and the Charing Cross Hospital on a single site. Such a decision could have a significant impact on two of Britain's leading biomedical research institutions, the Royal Postgraduate Medical School (RPMS) and the Medical Research Council's (MRC's) new Clinical Sciences Centre, both of which are currently based at Hammersmith (see *Nature*, 368, 176; 1993).

Mrs Virginia Bottomley, the secretary of health, announced on Monday that she had approved plans to combine the two hospitals under a single administrative trust, to be chaired by Sir Christopher Bland, the current chairman of the Hammersmith and Queen Charlotte's Hospital Special Health Authority.

Bottomley said that the joint trust, which will be responsible for working out a development plan for the two institutions, offered "the best way to enhance the qualities and reputation of its constituent hospitals." But she also said that it was likely to continue to operate from

its main sites "for the foreseeable future".

The decision was warmly welcomed by both scientists and administrators at the Hammersmith Hospital, who had been fighting a vigorous campaign to prevent being moved to the site of the Charing Cross Hospital. The dean of the RPMS, Sir Colin Dollery, said that the two hospitals would now "begin to examine the options for enhancing treatment and research on both sites."

David Evered, the second secretary of the MRC, was more cautious. While describing the trust as "an important step towards the future" of the hospitals, he emphasized that "full provision [is needed] from the outset for the necessary capital sums for facilities on both sites."

The government's announcement provided little indication as to whether such sums are likely to be made available. Indeed some fear it will have done little to lift the planning blight likely to remain hanging over research programmes at the Hammersmith until its future has been definitively decided. □

UK under fire over 'scrutiny' of government laboratories

London. The president of Britain's Royal Society has strongly criticized the government for its failure to consult more fully with the scientific community over potential changes in the way many government research laboratories are run.

The society issued a statement last week pointing to a "striking" contrast between the openness of discussions leading up to last year's white paper on science, and the conduct of a so-called 'scrutiny exercise' being used to pave the way for the privatization of a number of government laboratories.

It also warned that any new arrangements for the management of research and the ownership of research institutions that seriously weakened peer review — an implicit reference to proposals for the new Engineering and Physical Sciences Research Council (see *Nature* 365, 88; 1994) — "would gravely damage the national capability for scientific research".

The scrutiny exercise is being carried out under the guidance of Sir Peter Levene, the head of the Cabinet Office's Efficiency Unit. It was initiated shortly before Christmas but only officially announced at the beginning of February. A team of civil servants has been taking a close look at the activities of 53 government laboratories, in particular to discover which could be effectively run by the private sector.

Included in the list are five institutes run by the Medical Research Council, including its virology, radiobiology and developmental biology units, the Science and Engineering Research Council's Rutherford Appleton and Daresbury Laboratories, and the National Physical Laboratory (Britain's equivalent of the US National Institute of Standards and Technology) currently run by the Department of Trade and Industry.

Under standard government procedures, the review team was given 90 days in which to complete a report, and a draft copy was due to be presented to the minister for science, William Waldegrave, last week. The government argues that many laboratories now run as part of the civil service would operate more efficiently if their management was handed over to the private sector.

Government officials point out that privatization will not necessarily be the answer for all the laboratories being scrutinized, and that some might retain an independent, semipublic status. Indeed, one idea being considered is that the agencies that remain in the public sector should be grouped under a single Civilian Research Agency, reporting to the Office of Science and Technology (OST) and operating along the same lines as the existing Defence Research Agency.

But the Royal Society is worried that any

significant shift towards private ownership of the laboratories could damage their commitment to long-term research and undermine the government's access to independent scientific advice on important policy matters. It is also concerned at the speed of the current exercise, and the relative secrecy in which it appears to have been carried out.

"We feel that a lot of potentially very important decisions are about to be taken without adequate involvement of people outside in the scientific community," Sir Michael Atiyah, the president of the society, said last week. "We are worried that things are going very fast, and without widespread consultation or adequate scientific input."

According to Atiyah, the society is not against change, but is keen that the main



Atiyah: anxious

focus should be on whether the laboratories are doing their job adequately. "These institutions should be looked at pragmatically," he says. "The government should not go in with any *a priori* view about whether it is a good thing to privatize or not."

His views are echoed by the Institute of Professionals, Managers and Specialists (IPMS), the main union representing scientific and technical staff at the laboratories now being scrutinized, which is keen that the views of its members be heard on the fate of the institutions they work for.

"Our prime objection is that the government appears to be starting by asking which laboratories can be privatized, and then going on to discuss what to do with the others," says Valerie Ellis, deputy general secretary of the IPMS. "We welcome the idea of an efficiency scrutiny, but it should be focusing on how government research establishments can best support the objectives of the white paper."

Responding to the Royal Society's criticisms, Waldegrave said last week that, once the review is completed, "I will publish it, and will listen to the views expressed before reaching a decision." In the meantime, he said, "if people wish to make points about the scrutiny exercise, they can do so to either Bill Stewart [the government's chief scientific adviser] or myself."

OST officials say that a three-month period for public comment should allay the concerns of those who feel their points of view are being overlooked. But many scientists remain nervous. **David Dickson**

Lords panel backs ethical barriers to biotech patents

London. A key committee of Britain's House of Lords has given its support to attempts by the European Parliament to expand the use of ethical criteria in deciding what biotechnology inventions can be patented.

But it has opposed moves to ban patents on genetic engineering techniques for altering germ-line cells. Both the European Commission and Parliament want to exclude patents on such techniques, which they describe as "contrary to the dignity of man".

Such a view is held particularly strongly in Germany. But the UK committee says that applying such a test to patents would be "very difficult to interpret", and that the questions raised should be debated in a broader medical and ethical context.

The report, published last week by the House of Lords Select Committee on the European Communities, comes as a six-year effort in Brussels to harmonize biotechnology patent legislation between the member states of the European Union may be drawing to a close. A new draft of a revised directive was approved by the Council of Ministers in January, and will shortly go to the Parliament for a second reading.

The Lords report provides a general endorsement of the approach embodied in the current draft — in particular, its attempts to meet some of the concerns expressed by animal rights groups over patents on transgenic animals — and will boost the chances of the draft being approved before the Parliament dissolves in May.

The Lords committee says that current regulations are insufficient to ensure that biotechnological inventions are applied in a socially acceptable way and proposes an explicit list of techniques excluded from patenting, rather than general principles that would then need to be interpreted.

The report has received a relatively warm reception from some critics of animal patents. "We welcome the report, and in particular its awareness that there is a serious opposition to patents on living organisms," says Peter Stevenson, political director of Compassion in World Farming.

But Nicholas Scott-Ram of the British Biotechnology Group, says that many companies have three reasons for expressing concern about strengthening the role of ethical arguments in patent law: that it could reduce the competitiveness of the European biotechnology industry compared to that in the United States, where no such restrictions exist; that it places an unfair burden on patent officers to make ethical judgements; and that it will be difficult to ensure a "level playing field" in Europe, because different countries are likely to interpret the directive with differing degrees of vigour. □

Space station heads for battle in Congress . . .

Washington. The US National Aeronautics and Space Administration (NASA) is due to announce today (24 March) that the 'conceptual design' of the international space station is complete. It will then start praying that a status report to be delivered to Congress next week will be sufficient to secure continued US funding for the project.

A two-day review meeting being held at the Johnson Space Center, Houston, Texas, this week between NASA, its international partners and its contractors is an important landmark in the redesign of the space station. The latest — and, NASA promises, last — redesign began hurriedly last autumn when the Clinton administration decided to bring Russia into the programme.

"This is where we move from concepts to hardware implementation," says Randy Brinkley, NASA space station programme manager. "It is by far the most important milestone in the programme since last

year's redesign."

NASA says that three-quarters of the new international space station will be made up of hardware already planned for the original Freedom station. The project now moves into a detailed design phase, due for completion in April 1995.

Government officials from the international partners — the United States, Europe, Canada, Japan and Russia — met last week in Paris for the first time to discuss how existing agreements need to be modified to allow Russia into the programme. At the end of the meeting, they issued a vague communique, promising further negotiations starting in April, but giving no fixed date for their completion.

Each party has its own doubts and fears about the proposed collaboration. But its fate ultimately rests with the US Congress, whose House of Representatives agreed to continue it by only a single vote last sum-

mer, and will return to the issue over the next few months.

Once again, its prospects are finely balanced. The space station is likely to lose support from members of Congress who feel that the Russian deal may fall apart, or believe that it dilutes the station's value, either as a science project or as a 'pork barrel'. But it will gain others prepared to back it as a centrepiece of President Bill Clinton's policy towards Russia (*Nature* 365, 681; 1993).

George Brown (Democrat, California), chairman of the House Science, Space and Technology Committee, and until recently a key station supporter, is threatening to withdraw his backing, on the grounds that the programme will put too much strain on NASA's depleted resources.

But some observers in Washington see this as a ploy to extract concessions from the administration elsewhere. Ultimately, they expect Brown to return to the fold, while other Democrats who have previously shown little interest in the programme may rally to the cause if the success of the president's foreign policy is thought to be at stake. It is Republican supporters of the station who are most likely to defect.

In the end, two factors unrelated to the space station will probably determine its fate: the fiscal mood of Congress, and the state of US-Russian relations when the vote comes round in June. "Basically it's too early to call," says John Logsdon of the Space Policy Institute at George Washington University in Washington DC. "All we can say for certain is there's going to be a fight."

Colin Macilwain

Russian scientists set up green policy group

Moscow. A potentially influential new voice in debates over the future of Russia's environment emerged on the political scene in Moscow last week when two dozen scientists and public interest group representatives announced the setting up of the Centre for Ecological Policy.

The board responsible for running the centre will be chaired by Alexei Yablokov, a corresponding member of the Russian Academy of Sciences who was previously President Boris Yeltsin's adviser on ecology and public health and is now counsellor to the Security Council of Russia.

The new centre will have two main goals. One is to influence government policy in the field of ecology by proposing novel solutions to urgent ecological problems, the second to supply the environmental movement with sound scientific information and advice.

Carl Levitin

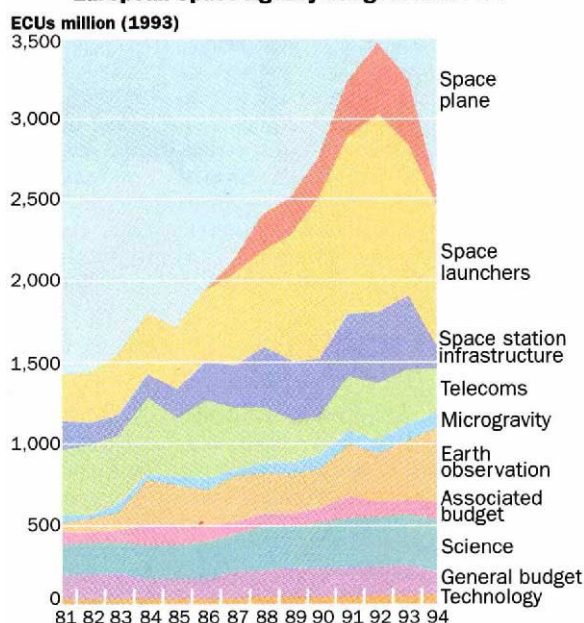
. . . as Europe cuts manned projects

Paris. Europe's decision on the terms of its participation in the international space station, which was to have been taken at a meeting of European Space Agency (ESA) ministers in February 1995, now seems likely to be postponed until after the French presidential election in May next year. It may even be delayed until after the German elections in October, as France and Germany are the largest contributors to ESA.

The 1994 budget of the ESA, which was agreed by the council in Paris last month, shows the full extent of the cut-backs in the manned space programme — including its commitment to the international space station — precipitated by continuing economic difficulties in its member states (see above).

ESA has abandoned the Hermès space plane, and plans to spend just ECU477.4 million (US\$529 million) over the next two years on studies of other programmes related to the space station (*Nature* 367, 305; 1994). These include the Columbus attached laboratory and an Apollo-like capsule to be carried on the launcher Ariane V. Moreover, no space station hardware will be built until ESA formally decides what

European Space Agency budgets 1981-94



type of contribution it is prepared to make to the station.

Science and Earth observation are the only two areas to receive more money in this year's budget. Other victims of the cuts include a flight of Europe's Spacelab module on the US shuttle, and a proposed mission of the Eureka unmanned retrievable instrument-carrying platform. The platform, which cost US\$400 million, flew in orbit only once, and will now be stored at a cost of \$8 million, as ESA cannot afford the ECU150 million needed to refurbish it for a second mission.

Declan Butler

NSF wins approval for 'new' research strategy

Washington. The US National Science Foundation (NSF) seems to have successfully headed off a threat by a Senate subcommittee to given some of its research funds to other agencies unless it switches more money from fundamental to applied research.

Last autumn, the appropriations subcommittee responsible for scrutinizing the NSF's budget told the agency, which pays for most non-biomedical research at US universities, that it would lose congressional favour unless it raised the proportion of research aimed at "strategic goals" to 60 per cent.

But at a hearing in Washington last week, Barbara Mikulski (Democrat, Maryland),

the chairwoman of the subcommittee, told Neal Lane, the director of the NSF, that the agency had already achieved "exactly what we wanted it to achieve". This was "an astounding accomplishment, knowing where you were last year", she said.

Lane told the subcommittee that he had set up a planning committee last month, and that this will report in May on how well the agency is meeting its strategic goals. A new financial tracking system is already in place, he said, and divisional directors are being encouraged to produce strategic plans similar to last autumn's Branscomb Committee report on high performance computing,

which Mikulski warmly praised.

"We're seeing an increased focus on basic math, physics and engineering research that addresses strategic goals," says Lane. This carefully chosen form of words appears to satisfy Mikulski — even though Lane's budget proposal for NSF contains none of the feared cuts in areas of basic scientific research.

Last autumn's tough language from the Senate subcommittee alarmed many basic researchers. But it appears to have been aimed less at NSF's basic research programme than at critics of the agency's bid for continued expansion at a time of fiscal constraint.

Both Mikulski's committee and its House counterpart will process the president's requested 6 per cent increase in NSF's budget during the summer. The final amount allocated to building new research laboratories is likely to be increased, with a corresponding reduction in extra money for research projects. There are already strong indications that the total Clinton budget for 1995 will be substantially reduced by Congress, affecting budget projections at the NSF and all other science-funding agencies.

Separately, the first NSF authorization bill since 1988 is likely to be finalized by the House Science Space and Technology (SST) committee this week. A Republican amendment that would bar the NSF from giving any grants to universities that accept earmarked funds from Congress is likely to fail, after committee chairman George Brown (California, Democrat) declined to support such a move.

Colin Macilwain

Senate boost for technology plans

Washington. The Clinton administration's ambitious plans to implement a wide-ranging technology policy received a boost last week when the Senate passed a bill authorizing a major expansion of the federal government's programmes to assist industry.

But the budgetary provisions of the National Competitiveness Act, much of which was originally drafted by Vice President Al Gore when he was still a senator, were scaled back by a third under pressure from Republican critics. The bill now proceeds to a conference with two related bills already passed by the House of Representatives, probably resulting in a final version for both houses to pass and the president to sign by the end of the year.

One of the most significant parts of the Senate bill expands programmes through which the government part-funds small companies to apply new technology or bring it to market. The focal point for this activity will be the Commerce Department's National Institute of Standards and Technology (NIST), although the Senate's version of the bill spreads some of the responsibility to other government departments.

Other sections of the bill authorize research to support the development of the so-called information superhighway. The amount of money to be spent on this is small compared to private sector investment in telecommunications. But Gore believes that it can play a crucial role in ensuring that the superhighway can be used by schools, hospitals, government departments and low-income households.

Although industry lobby groups support the programmes designed to help small businesses, Republican opponents of the bill remain sceptical about the way in which Ron Brown, Commerce Secretary and former Democratic Party chairman, will use the money. Despite the due processes to which NIST must adhere in allocating the funds, they are concerned that money will be used

to shore up Democratic support in key states such as California.

But, wary of recent Clinton attacks on them for resisting everything he proposed in Congress, the Republicans stepped back from talking out the bill, and struck a deal allowing it to pass with a total two-year funding cut from \$2.8 billion to \$1.9 billion.

Another contentious issue that remains to be resolved between the House and the Senate is the so-called Manton amendment to the House version, which excludes companies with foreign directors from participating in any of the technology schemes. The Senate is prepared to unite with the administration and industry lobby groups to try to remove the Manton amendment from the bill.

Colin Macilwain

Women in industry to get a helping hand

San Francisco. A programme launched three years ago in US universities to support younger women scientists by pairing them up with a more experienced 'mentor' is about to be extended to scientists working in industry.

The programme is run by the Association for Women in Science (AWIS), and has been developed with success in universities across the United States, with 32 separate chapters participating in 1993.

The Palo Alto chapter of AWIS already has been working with a few women who work in industry to encourage a greater business perspective among science students at Stanford University, the University of California at Berkeley and at San Francisco.

The chapter now plans to extend the mentors' services to women scientists working in local biotechnology and high technology companies. The aim is to help companies both retain key scientists and develop their potential, says Sherrie

Wilkins, president of the chapter.

Shirley Johnson, head of a research and development group at Biocircuits Corp. in Sunnyvale and a co-director of the AWIS mentoring programme this year, says that relationships established through the programme provide an important form of support to women scientists at all levels.

Johnson says one of the aims of AWIS is to increase the number of women who are successful in scientific careers. At present, only 36 per cent of women who teach science or engineering at four-year colleges and universities hold tenured positions, compared with 59 per cent of men.

The national AWIS mentoring programme over the past three years helped pairs of women develop strong relationships that would guide a less-experienced partner through practical matters such as how to work with a principal investigator on a project, or how to look for a job.

Sally Lehrmann

Canadian AIDS suit raises hopes for HIV-blood victims

Montreal. A flood of lawsuits is expected to follow the decision to award more than half a million Canadian dollars to the HIV-infected wife of a man who died of AIDS after receiving a transfusion of HIV-contaminated blood. Kenneth Arenson, the lawyer for Rochelle Pittman, said last week that he has 19 other clients who intend to pursue legal action against the Canadian Red Cross Society, individual hospitals and the province of Ontario.

The outcome of the Pittman case was announced on the eve of a deadline set by all except one of Canada's provinces and territories for acceptance of a compensation package by more than a thousand recipients of HIV-infected blood. To obtain the money, victims had to sign waivers promising not to sue Ottawa, the provinces or territories, hospitals, the Red Cross and pharmaceutical and insurance companies.

But many consider the package inadequate, as it offers far less than the C\$515,000 (US\$380,000), plus interest and legal fees, that Pittman — who was ineligible for the package because she was infected sexually through her husband, and not directly — is likely to receive.

The package also less generous than the province of Nova Scotia, which, as well as making annual payments of C\$30,000 to victims, pays for expensive AIDS drugs, as well as funeral costs and post-secondary education for their children.

The Ontario Court ruled in Pittman's case that the Red Cross, the Toronto Hospital and her general practitioner were negligent in failing to tell her husband that he had been infected by contaminated blood. His wife discovered later that she had become HIV-positive.

But the court ruled that the Red Cross had not been negligent in its efforts to safeguard blood supplies before 1985, when national screening for HIV started. This will complicate the legal claims of other haemophiliacs and transfusion patients.

Pittman's was one of many stories presented to a commission of inquiry into Canada's blood supply system, set up last October. The commission is holding nationwide hearings on the working of the system, including the events surrounding contamination of blood supplies in the early 1980s. Its report is expected to be completed by the end of September.

More than two million Canadians received transfusions between 1978 and 1985, and epidemiologists believe that, in addition to the thousand or so individuals known to have been infected in this way, a further 200 may also have been infected.

David Spurgeon

French fast reactor proposal still breeds controversy

Paris. Last month's decision by the French government to restart the Superphénix fast-breeder reactor at Creys-Malville near Lyons, but to shift its main purpose from generating electricity to research into incinerating plutonium, has done little to dampen the controversy that continues to surround the plant.

Superphénix, a 1,200-MW prototype reactor, was closed in 1987 following repeated leaks in the liquid-sodium cooling circuits. It was to be the first of a new generation of fast breeders, designed to breed plutonium from stocks produced by France's reprocessing plant at la Hague.

The strategy anticipated that a spread of conventional reactors would lead to a shortage of uranium and make plutonium the fuel of choice. But this has not happened, and uranium remains plentiful. Moreover, such logic has been eclipsed by growing concern over the safety of storing and transporting more than 1,000 tonnes of plutonium stockpiled around the world.

Under the new plan, NERSA — a consortium of French, German and Italian utilities that owns the reactor — and the French Atomic Energy Commission (CEA) will commence a research programme aimed at converting Superphénix into a prototype for a purpose-built plutonium furnace.

The programme would be based on the CAPRA programme (Consommation Accrue de Plutonium dans les Rapides), which CEA began last autumn (see *Nature* 365, 381; 1993). This proposes equipping Superphénix with cores lacking a uranium blanket in 1995 and 1999 — converting a fast neutron reactor from a breeder to a consumer — and trying to raise by a factor of three the estimated maximum rate of plutonium consumption.

The reactor may be started up at low power as soon as this summer, once the proposed programme has been reviewed by Robert Dautray, the head of CEA, and Claude Detraz, director of the Institute of Nuclear and Particle Physics at the Centre National de la Recherche Scientifique (CNRS).

But many are still sceptical of the wisdom of restarting the reactor. The decision is widely seen as a political solution to the problem of satisfying the utilities companies, who have paid FF27.7 billion to build and more than FF600 million annually to run a reactor that has operated for a total of less than six months.

In particular, many feel that as Superphénix was not designed as a research reactor, it can only be an extremely expensive research tool poorly adapted to the job. Georges Vendryes, a former director of CEA

and 'father' of the reactor, has admitted that "we can't say that this is the ideal reactor for this sort of work".

Others doubt the technical feasibility of incinerating plutonium. Britain rejected the idea in the late 1970s on safety

and economic grounds. One senior scientist who worked on the European fast reactor programme says that incineration would be incomplete, and would leave unacceptably high levels of plutonium and minor actinides.

He adds that the utility companies will not pay for the new fast breeders and extra processing that large-scale incineration would require. "It's not a solution," he says.

Despite such criticisms, Electricité de France (EDF) has welcomed the proposal, and says it hopes that research using Superphénix will lead to France building one plutonium-furnace for every five water-pressurized reactors.

Furthermore, Gerhart Heusener, who heads nuclear safety at the Karlsruhe Nuclear Research Centre in Germany and leads German participation in CAPRA, has also argued that the major stumbling block to incineration is simply lack of research. "Perhaps incineration is no use," he says. "But CAPRA will at least let us find out."

EDF's partners in NERSA say that any option would be better than closing Superphénix — which would have cost around FF18 billion on top of their original investments. But they are waiting until they have a meeting with the French government before taking a formal position.

Alexandre Autieri, the deputy permanent secretary of NERSA, says that "we will try to use Superphénix as a research reactor, but our major interest is to produce electricity". He points out that a fast-breeder produces as much electricity when it consumes plutonium as it does when it breeds it, and says "the new orientation is possibly not opposed to our original aim".

Indeed, although one result of the switch in the goals of Superphénix is that EDF will no longer be able to dictate power levels, the output of the reactor may be progressively increased up to much the same level as that originally intended.

Declan Butler

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Superphénix — back in business.

Marc Moreau/La Photographie EDF

Japanese telephone shares finance four new institutes

Tokyo. An organization set up in 1985 with proceeds from the privatization of the Nippon Telegraph and Telephone Corporation (NTT) — Japan's main domestic telephone company — has announced plans to provide billions of yen over the next five to seven years to support three new institutes that are being established to develop computer network and mobile telecommunications technology.

The Japan Key Technology Centre, which already funds several wealthy semi-private institutes in Japan, also plans to make a large investment in a laboratory that will develop optical techniques for detecting and analysing trace amounts of biomolecules, such as messenger RNA. The ultimate goal of the laboratory is to develop techniques for looking at processes such as gene expression as they occur in living organisms.

The centre is affiliated to the Ministry of Posts and Telecommunications (MPT) and the Ministry of International Trade and Industry (MITI). It receives between ¥25 and ¥28 billion (US\$238 to \$267 billion) each year from the dividends of NTT shares held by the government, most of which is subsequently invested in research institutes that are supported and run by consortia of private companies.

The best known (and best endowed) of these are the Protein Engineering Research Institute in Osaka, and the Advanced Telecommunications Research Institute International in the emerging Kansai science city which is being built between Osaka, Nara and Kyoto. Both have won international recognition for the high quality of their research (see *Nature* 359, 577; 1992).

Last week, the centre announced the establishment of five new institutes. In line

with the government's recent investments in computer networks, one institute (which will be supported by Fujitsu, NEC and Hitachi) will develop technology for high speed computer networks with transmission rates up to 2.4 gigabits per second, 400 times the present fastest speed of networks in Japan.

Under the centre's standard procedures, it will provide 70 per cent of the institute's ¥3.5 billion annual budget over the next five years. The companies will provide the rest of the money.

Two further institutes will support work on the development of advanced mobile telecommunications systems. One will focus on satellite technology, and the other mainly on land-based systems, including the application of high-temperature superconductors to such systems.

The most novel institute is the Laboratory of Molecular BioPhotonics, headed by the small company Hamamatsu Photonics. The company makes some of the world's most sensitive photon detecting devices, and has been associated with a number of government-funded science projects. The company's giant photomultiplier tubes, for example, are used in the Kamiokande neutrino detector in a mine in Gifu Prefecture (see *Nature* 365, 596; 1993).

The biophotonics laboratory will open towards the end of the year in a science park in the town of Hamamatsu, between Tokyo and Nagoya, where the company is located. It will develop fluorescent reagents for labelling functional biomolecules, such as messenger RNA, as well as highly sensitive optical detectors that can be used to pinpoint trace quantities of the molecules.

David Swinbanks

Final agreement on new funding for Framework projects

Paris. The next five-year Framework research programme of the European Union (EU) overcame its last major obstacle on Monday when the Council of Ministers and representatives of the European Parliament reached agreement on a single text.

This is the first time that the conciliation procedures between the two political bodies, which were introduced under the terms of the Maastricht agreement, have been tested.

Last autumn, the parliament asked for a budget of ECU13.7 billion (\$US15.2 billion) for the Framework programme for the period 1994-98 inclusive. In contrast, however, various member states, including both Britain and Germany, argued for a figure as low as ECU8 billion.

In December, the Council of Ministers agreed on a proposal that would provide an immediate commitment to spend ECU12 billion, while 'freezing' an extra ECU1 billion. But last month, parliament rejected this solution and argued for a definite commitment to an extra ECU400 million.

The compromise reached on Monday means that the Fourth Framework will definitely receive ECU12.3 billion, with a decision on an extra 0.7 billion to be taken in 1996. The enlargement of the EU to include Finland, Sweden and Austria will swell the Framework budget by an extra 6 per cent.

The parliament also won a commitment from the council that a significant proportion of the extra ECU300 million will be spent on technology transfer — the application of research results. Despite growing calls for the EU to spend more on this activity, member states had previously cut the money proposed for it from 4.6 per cent of the total, as suggested by the European Commission, to only 2.5 per cent.

The conciliation committee also agreed a figure of ECU900 million for the EU's four Joint Research Centres (JRCs). Parliament had proposed ECU910 million for the JRCs, and the council ECU885 million. Parliament refused to accept that the centres should be made to compete for 22 per cent of their budget, even though this was reduced from the council's original suggestion of 24 per cent.

But it is now virtually certain that the Framework programme will be adopted quickly. Under the new procedures of the Maastricht agreement, both council and parliament must approve a joint text of the Framework budget.

There had been widespread fears that, if the complex approval procedure was not completed before June, when elections to the parliament are due to be held, it would have to begin again in 1995.

Declan Butler

New institutes of Japan Key Technology Centre

Institute name*	Budget (billion ¥)	Duration years	Participating companies
Laboratory of Molecular BioPhotonics	6.4	6	Hamamatsu Photonics; Chugai Pharmaceutical; Toyota; Japan Tobacco; Nikon
Commuter Helicopter Advanced Technology Institute	8.8	7	Kawasaki Heavy Industries; Shimadzu; Teijin; + 2 others
High-speed Network Computer Technology Institute	3.5	5	NEC; Fujitsu; Hitachi
Next-generation Satellite Telecommunications and Broadcasting Systems Institute	8.4	7	NEC; Mitsubishi Electric; Hitachi; Fujitsu; KDD; Pioneer; Japan Radio; + 6 others
Mobile Telecommunications Advanced Technology Institute	2.9	6	Nippondenso; Alps Electric; + 1 other

*Translation of tentative name in Japanese

Ecologists clash over 'too academic' research

Basel. The Swiss National Science Foundation (NF) is planning to set up a new programme on applied biodiversity research, following criticism by ecologists working for non-academic organizations that too much money is going to academic projects with little immediate relevance to conservation.

The Swiss government originally announced in 1991 the launch of a biodiversity research programme costing SFr5 million (US\$3.5 million) for the three years 1993 to 1995 as part of its preparations for the 1992 Earth Summit in Rio de Janeiro.

By the spring of 1992, the NF had received almost 190 grant applications, most coming from independent ecologists or environmentalists working for Switzerland's many small conservation agencies. But only two of the 20 grants finally approved were given to researchers outside universities or academic research institutes.

Non-academic ecologists believe their chances of receiving a fair share of the funds were reduced by the fact that they were not represented on the 16-strong grant committee (though it included two industrialists).

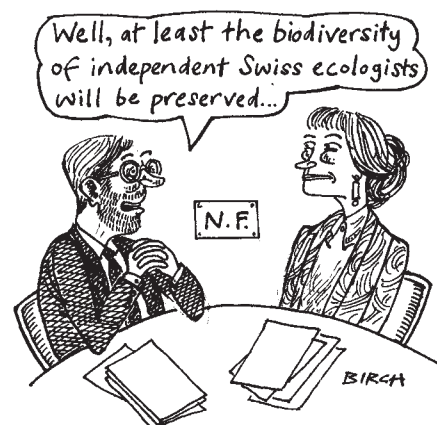
The distribution of the funds also gave rise to charges of conflict of interest, as many of the grant recipients were academic colleagues known to the committee members. Rudolf Häberli, the NF official responsible for the biodiversity programme, denies that there was any wrongdoing, pointing out the difficulty in a country with a small scientific community of identifying scientists with no connection to each other.

Furthermore, one of the committee members who received a grant award says that he left the room when his application was being discussed. Nevertheless, sensitive to the criticisms, the NF has now changed its internal regulations covering participation

in expert committees.

Häberli also denies that the allocation of research funding was unfairly biased towards university-based groups. He emphasizes that the applications were reviewed by international experts, and judged only on their scientific quality.

But Hans-Jörg Blankenhorn of the Swiss Game Directory, the body responsible for regulating game management and hunting regulations, is one of the critics who claim that the money should have been used more



Slovak university reforms stall

Bratislava. The collapse of the Slovak government last week seems likely to bring new delays to university reforms designed to compensate for shortcomings in legislation introduced after the collapse of communism four years ago.

The reforms have already been stalled by a combination of the low priority given to academic issues by a succession of unstable governments and disagreement within the academic community over some of the proposed changes.

Czechoslovakia was the first Central European state to pass a new higher education law after the fall of communism. The 1990 act, which was intended to introduce Western-style values into the universities, is still valid in both of the independent republics into which the country split.

The new law set up an accreditation committee to set academic standards and decide the status of each of the 57 faculties in Slovakia's 14 universities. But the high level of autonomy given to universities in reaction to the external interference of communist times has had its price.

Each faculty is now independent, and there is very little inter-faculty cooperation. This means that university rectors have little power to coordinate the efforts of their universities as a whole, and therefore to carry out broad reforms.

Last September, the Ministry of Education, Youth and Sports set up an advisory committee to propose changes in the 1990 act that would make it more workable. But the committee's proposals have not been well received.

One point of controversy is the suggestion that rectors should be appointed as the head of university senates, and given the

power to overrule decisions made by individual faculties. "This is how things work in Western European countries, and it is something that the committee is not prepared to compromise on," says Juraj Sinay, a committee member and pro-rector of the Technical University in Kosice, who claims that parliament agrees with this position.

Equally controversial is the issue of who should have the right to award PhDs. This time the opponents are the universities and the Slovak Academy of Sciences. In communist times, both gave research degrees, then known as candidate of science degrees.

But Miroslav Urban, a member of the accreditation committee, argues that with the recent establishment of Western-style PhD programmes, only universities should have the right to award such degrees.

Unsurprisingly, the academy is not happy with this arrangement. It has had the right to award doctorates since 1953, and is not ready to give this up without a struggle. But a compromise may be possible. Ján Kazár, scientific secretary of the academy, says its minimum requirement is "to participate in preparing colleagues for a scientific career".

If the reforms are passed by parliament in their present form, all university staff will face an evaluation of their competence, and their posts will become open to competition.

The reform committee is keen for the bill to be approved by parliament before the start of the autumn term, arguing that the apparent democracy of "faculty power" has been a block to true reform, as faculties have proved resistant to change. But the fall of the Slovak government makes this unlikely — even if the academic community can decide what it really wants.

Alison Abbott

directly to tackle environmental problems.

He criticizes, for example, a SFr4 million grant to scientists at Basel University to support basic studies on the effect of habitat fragmentation on the genetics of populations, and the effect of increased carbon dioxide levels on grassland plant communities. But Bruno Baur, a zoologist who leads one of these projects, defends his research on the grounds that it does contribute to conservation, even if only in the longer term — a view officially supported by the NF.

Such arguments have failed to appease the critics, who argue that much more basic information is needed on the flora and fauna of Switzerland to help protect them, and that this is where more of the NF money should have gone. They also claim that some academic projects are being re-labelled as biodiversity projects merely to qualify for the new funds.

"Universities still look down on what we are doing", says Gabi Gerlach of the consultants BiCon AG in Kreuzlingen, one of the ecologists whose grant applications were rejected by the foundation.

The NF appears to be responding to such complaints. Last November, individuals whose proposals had been rejected were invited to take part in discussions on a new special programme of applied biodiversity research. The NF will decide in May whether to go ahead with the programme. If it gets government approval, the programme will be submitted to parliament for ratification in the summer of next year, and will be ready to start in 1996.

Oliver Klaffke

Elementary errors about entropy

SIR — We wish to correct a number of gross if elementary errors in John Maddox's article "When entropy does not seem extensive"¹.

First, entropy increases with increasing disorder, and does not decrease, as he claims. Second, it is the Helmholtz free energy and not the entropy difference that is proportional to $-\log Z^n$, where Z is the molecular partition function and n the number of molecules.

Maddox is also wrong to claim that the entropy of a black hole can be "equated with the calculable entropy of that radiation [from the black hole], whence the result that the entropy is proportional to the surface area". Even black-hole specialists allow that the radiation is black-body radiation, for which the entropy is proportional to the volume and the cube of the temperature. We know of no radiation whose entropy is proportional to the surface area.

That is one reason why it seems to us bizarre that the entropy of a black hole should be taken to be proportional to its surface area. The fact that the heat capacity of a black hole is negative is also a reason why such an object can never come into equilibrium with its radiation (among other things this suggests the Carnot rule does not apply, as Maddox asserts).

In our opinion, the extensivity property of the entropy is a red herring. The characterizing property of any entropy is its concavity^{2,3}, which ensures that heat should not spontaneously flow from cold to hot, that pressure should not increase with volume and that the chemical potential should not decrease as the particle number increases. The Beckenstein-Hawking entropy lacks concavity and so violates the second law.

B. H. Lavenda

Dipartimento di Scienze Chimiche,
University of Camerino,
Italy

J. Dunning-Davies

Department of Applied Mathematics,
University of Hull,
Hull HU6 7RX, UK

1. *Nature* **356**, 103 (1993).

2. Lavenda, B. H. & Dunning-Davies, J. (*Found. Phys. Lett.* **3**, 435–441 (1990)).

3. Lavenda, B. H. *Statistical Physics* (Wiley, New York, 1991).

Baha'i religion

SIR — Readers of *Nature* may be interested to know about the Baha'i religion, founded by Baha'u'llah in the mid-nineteenth century, which advocates science and religion as the two

potent forces in human life.

Baha'u'llah's teachings place great emphasis on spiritual principles as major contributors to all human achievements and hold religion to be the most effective means of establishing order in society. They call for recognition of the oneness of the human race as the prerequisite for the establishment of a world civilization. As incidentals to this central principle, Baha'u'llah teaches the abandonment of prejudices and blind imitations and calls for independent investigation of truth, universal education, equality of the sexes, essential harmony of science and religion and the need for a universal language. He also lays the blueprint for the future world civilization.

Commenting on some features of this world civilization envisaged by Baha'u'llah, Shoghi Effendi, the Guardian of the Baha'i religion, wrote in 1938:

"National rivalries, hatreds, and intrigues will cease, and racial animosity and prejudice will be replaced by racial amity, understanding and cooperation. The causes of religious strife will be permanently removed, economic barriers and restrictions will be completely abolished, and inordinate distinction between classes will be obliterated. Destitution on the one hand, and gross accumulation of ownership on the other, will disappear. The enormous energy dissipated and wasted on war, whether economic or political, will be consecrated to such ends as will extend the range of human inventions, and technical development, to the increase of the productivity of mankind, to the extermination of disease, to the extension of scientific research, to the raising of standard of physical health, the sharpening and refinement of human brain, to the exploitation of the unused and unsuspected resources of the planet, to the prolongation of human life, and to the furtherance of any other agency that can stimulate the intellectual, the moral and spiritual life of the entire human race."¹

Keyvan Nazerian

Avian Disease & Oncology Laboratory,
USDA/ARS,
East Lansing,
Michigan 48823, USA

1. Shoghi Effendi, *World Order of Baha'u'llah*, 204 (Baha'i Publishing Trust, Wilmette, Illinois, 1982).

Epidemiology

SIR — Williams *et al.*¹ propose that terms applied to human disease should not be used for animal diseases. We dispute the suggestion that the terms describing the occurrence of diseases in populations should be distinguished by their hosts of principal significance.

We recognize that *-dem-* derives from the Greek *demos* (people) but we are sure that the authors of the letter would be comfortable with the term 'population dynamics' despite the fact that 'population' derives from the Latin '*populus*'.

Thus the only justification for the use of 'epizootic' seems to be that it is more specific than the generally accepted term 'epidemic'. The logical extension of this is that the authors of the letter in question (many involved in fisheries research) should be promoting the term epichthyotic while the study of disease in populations of birds should become epiornithology. This would lead to a ridiculous and confusing proliferation of unnecessary terms.

Many diseases are common to humans and nonhuman animals. Epidemiology, as a holistic discipline, should not create artificial distinctions. Would the authors prefer an individual investigating the natural history of toxoplasmosis to describe the study of its transmission among cats and sheep as epizootiology, that of its behaviour in humans as epidemiology and perhaps the transmission from animals to man as epizoodemiology?

The fact that the term epidemiology is both preferred and accepted for the study of diseases in populations in animals can be seen from the fact that the four major general textbooks published in the discipline^{2–5} all have the term epidemiology in the title. The primary journal for veterinary epidemiological studies (*Preventive Veterinary Medicine*) also encourages the use of epidemiology. We suggest that the terms epizootic and epizootiology be completely removed from the scientific literature.

Ian R. Dohoo (*Epidemiology, Atlantic Veterinary College, University of PEI, Charlottetown, PEI C1A 4P3 Canada*);

Roger S. Morris, (*Massey University, Palmerston North, New Zealand*); **S. Wayne Martin** (*Department of Population Medicine, Ontario Veterinary College, University of Guelph, Ontario N1G 2W1, Canada*); **Brian D. Perry** (*Epidemiology and Socioeconomics, International Laboratory for Research in Animal Diseases, PO Box 30709, Nairobi, Kenya*); **Theresa Bernardo** (*Agriculture Health Program, Inter-American Institute for Cooperation on Agriculture, PO Box 2200–55 Coronado, Costa Rica*); **Hollis Erb**

(*Epidemiology, NY State College of Veterinary Medicine, Cornell University, NY 14853–6401, USA*); **Michael Thrusfield** (*Department of Veterinary Clinical Studies, Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK*); **Ron Smith**, (*Division of Epidemiology and Preventive Medicine, College of Veterinary Medicine, University of Illinois, Illinois 61801, USA*); **Valdir R. Welte** (*Animal Production and Health Division, Food and Agriculture Organization, 00100 Rome, Italy*)

1. Williams, E. H. *et al.* *Nature* **364**, 664 (1993).

2. Schwabe, C. W., Riemann, H. P. & Franti, C. E. *Epidemiology in Veterinary Practice* (Lea & Febiger, Philadelphia, 1977).

3. Thrusfield, M. *Veterinary Epidemiology* (Butterworth, London, 1986).

4. Martin, S. W., Meek, A. H. & Willeberg, P. *Veterinary Epidemiology: Principles and Methods* (Iowa State Press, Ames, 1987).

5. Smith, R. *Veterinary Clinical Epidemiology. A Problem-Oriented Approach* (Butterworth-Heinemann, Boston, 1991).

How convincing is DNA evidence?

David J. Balding and Peter Donnelly

Although some of the initial controversy surrounding DNA profiling has been resolved, courts have been misled about the strength of DNA evidence.

DNA "fingerprinting" was widely hailed as the greatest breakthrough in forensic science this century. Its ability to "establish guilt or innocence" captured the attention and enthusiasm of police, politicians, the legal establishment and the general public. Not for the first time, initial enthusiasm for a scientific panacea has since been somewhat tempered.

In the first UK challenge to a conviction based on DNA evidence, the Court of Appeal recently quashed the conviction of Andrew Deen for rape, declaring the original verdict to be unsafe. In ordering a retrial, the Lord Chief Justice expressed a desire not to undermine all use of DNA profiling for forensic identification. Nonetheless, the judgement highlights new concern about the use of such evidence.

Since becoming widely used in the late 1980s, forensic applications of DNA profiling have attracted controversy. Initially, the reliability of laboratory procedures and the validity of assumptions about population genetics that underlie the statistical calculations were questioned^{1,2}. Tempers, and media interest, were raised by accusations of "dirty tricks": improper pressure on journal editors, leaking of confidential referees' reports, harassment of defence experts and undisclosed conflicts of interest³⁻⁵.

In 1990, the National Research Council (NRC) of the US National Academy of Sciences instituted a panel to consider the scientific issues. Rather than settling the debate, its report⁶ caused further controversy. Although many of its recommendations have been implemented in US courts, concern that the report leans too heavily in favour of the defence has recently led to pressure from the FBI and others for a new study. The manner of this intervention has itself provoked criticism⁷.

What is the basis of the technique that has generated all this controversy? DNA profiling is based on the pattern of measured lengths of certain DNA fragments, which appear as 'bands' on autoradiographs. In forensic identification, the DNA profile obtained from a sample of hair, blood or semen found at the scene of a crime is compared with that of a suspect. If discrepancies between the two observed profiles cannot be attributed to measurement error, the suspect will usually be exonerated. This ability to establish innocence is one of the great benefits of the technique. The London Metropolitan

Police Forensic Science Laboratory, for example, has found that about 20 per cent of suspects in rape cases are excluded by DNA profiling (M. Greenhalgh, personal communication).

If two DNA profiles are judged to match, the strength of this evidence is usually measured by a "match probability" — the probability that an unrelated individual chosen randomly from an appropriate population will match the crime profile. Much controversy has centred on methods for calculating this match probability, but little attention has been directed at the more fundamental issue of its correct interpretation. There are two important questions in relation to interpretation of DNA evidence. First, what is the probability that an individual will match, given that they are innocent; and second, what is the probability that an individual is innocent, given that they match? The match probability is an answer to the first question, but it is the second which is of direct interest to courts.

The two questions are quite different. A common error, the 'prosecutor's fallacy', consists of giving the answer to the first question in response to the second, which amounts to confusing the match probability with the probability of innocence. In the original trial of Deen, for example, the forensic scientist described his match probability as the "probability of the semen having originated from someone other than Andrew Deen". He went on to agree that "the likelihood of this [the source of the semen] being any other man but Andrew Deen is 1 in 3 million". Each of these statements is an example of the prosecutor's fallacy, and this misrepresentation of the DNA evidence was one of the two reasons given in the judgement for the success of the appeal. (The other related to the judge's summing up of a dispute over possible non-matching bands.)

It is important to appreciate that a small match probability does not, in itself, establish guilt. Consider a hypothetical case, with match probability 1 in 1 million, in which there is little or no evidence other than the DNA profile match. (Because of its perceived strength, cases involving DNA evidence are often brought to court with little additional incriminating evidence.) There may be many individuals who, if not for the DNA evidence, would be as likely to be the culprit as the defendant. If there were 500,000 such

individuals, for example in a large city, the probability that the defendant is innocent after taking the DNA evidence into account would be at least one-third*. It would thus be extremely prejudicial in this case to tell the jury that the chance that the defendant is innocent is 1 in 1 million. Intuitively, if the pool of possible perpetrators is large enough to ensure a non-negligible chance that at least one other individual has the profile, then even with a very small match probability the DNA evidence may not be overwhelming.

In Dean's appeal, the judges were explicit that it was not necessary for them to assess how much difference the correct presentation of the DNA evidence would have made to the outcome. How many other British appeals might succeed on this ground alone? The prosecutor's fallacy in the context of DNA also appears to be widespread in the United States⁸. It will be interesting to see how courts in other countries rule on appeal cases.

Although the interrogative nature of the adversarial system can hamper accuracy of expression, the argument that the jury would have reached the same decision even if the DNA evidence had been properly presented raises two serious concerns. Some of the infamous miscarriages of justice recently exposed in the United Kingdom involved scientists obscuring details of their evidence. Perhaps equally disturbing is the possibility that the jury would have made the same decision in any event. Particularly in view of exaggerated media claims, jurors may interpret the very small probabilities that arise in connection with DNA evidence as overwhelming proof of guilt, irrespective of the other evidence, whether or not the correct wording is used. Those adducing and presenting such evidence must ensure that the danger of its misinterpretation is carefully explained to the jury.

The issue of whether the defendant is the source of a crime sample depends on

*The basis for this probability calculation is as follows. There are two possibilities: either the defendant is the criminal, in which case a match is certain; or one of the other 500,000 individuals is the criminal and they match by chance. The former has probability $(500,001)^{-1} \times 1$; the latter has probability $500,000 \times (500,001)^{-1} \times 1,000,000^{-1}$. Given that one of the two possibilities occurred, the probability that it is the second (innocence) is thus $\frac{1}{3}$. If there are in fact other possible perpetrators, the probability of innocence will be increased.

the DNA evidence and the other evidence in the case. The assessment of the other evidence is a matter for the jury, not an expert witness. Consequently it is inappropriate for an expert to express a view on whether the defendant was the source of the crime sample or on the guilt of a defendant. Although it may be natural for scientists to think of their role as performing a test to decide whether the defendant is the source of the crime sample, this view is misguided. (An analogy with dermal fingerprints is inaccurate, because the uniqueness of dermal fingerprints is well-established in the courts.)

Even when the match probability is small, its exact value can be important. One common response to concern about assumptions underlying the calculation of the match probability is that they will not affect the probability in a way that matters "in practical casework". Some take an even more extreme view. Philip Reilly, a member of the NRC panel, has said⁹ that "there is absolutely no need to come in with statements like 'one in a billion', 'one in 10,000 is just as good'". Such views are incorrect. As the hypothetical example above illustrates, changes of one or two orders of magnitude in match probability can mean the difference between conviction and acquittal. Whether this applies in a particular case depends on the jury's assessment of the other evidence, and is not in the domain of the scientist.

Another important general point, often overlooked, is that the probability that a particular individual has the profile in question is, in general, different from the probability that he has this profile given that the defendant has it. The latter probability is the relevant one in forensic applications; the difference between the two can be substantial. For a four-locus profile with (unconditional) match probability of 10^{-7} , the conditional match probability for a sibling of the defendant will usually be greater than 10^{-2} , while that for another member of the defendant's subpopulation might be 10^{-5} or more.

The current practice of ignoring close relatives, unless there are good reasons to suspect them, will often greatly overstate the weight of the DNA evidence. In a recent Scottish case (*HMA v. Aslam*) the forensic scientist reported a match probability based on three single-locus probes, of 1 in 49,000 for unrelated individuals, adding that a relative "will be more likely to have the same DNA profile". He accepted that the probability of a match from a particular brother of the defendant was about 1 in 16. As it happened, the defendant had five brothers. If the other evidence did not distinguish between the six brothers, the probability of the defendant's innocence would be more than 1 in 5; even if there had been only one brother, this probability would have been at least 1 in 17.

Because human populations are not homogeneous, an available database may not provide reliable information about allele frequencies in ethnic groups relevant to a particular crime. This has been the focus of much debate. A related issue results from positive correlations in genetic traits due to shared ancestry on an evolutionary timescale. Knowledge that the defendant has the profile will increase the probability that other unrelated individuals in the same (and possibly other) subpopulations will also have the profile.

Not only will match probabilities differ between different individuals, but in many cases the other evidence will make some possible perpetrators more likely than others to be the true culprit. In principle, the jury must assess both the match probability and the other evidence, for every possible perpetrator, in coming to a verdict. The task of explaining this procedure to a jury seems daunting, particularly in an adversarial legal system in which obfuscation may serve the interests of one side.

A careful theoretical analysis does, however, show that some concerns about DNA profiling are misplaced. These include the choice of reference population and the appropriate statistical treatment when the defendant has been identified from a search through a DNA profile database. There remains a danger that subtle statistical issues may be overlooked in unusual cases. Examples include cases in which experimental 'artefacts' are invoked to explain inconvenient features on autoradiographs, those in which a match is declared despite bands falling outside match 'guidelines', and those in which the most convincing of several replicated profiles is presented in evidence. For example, at various stages in the Deen trial and appeal, different explanations were suggested for bands that appeared in one profile but not the other. Merely positing an experimental explanation is not sufficient; a different statistical analysis is also necessary.

Although progress has been made on laboratory standards, forensic scientists are still being criticized for "secrecy and self serving bias" in the performance and reporting of proficiency tests (see ref. 10). When extremely small match probabilities are claimed, it seems naive, at best, to ignore the possibility of false-positive results through human error. In a recent case (*R. v. Clapham*), a forensic scientist described his laboratory's high standards of internal quality control. On discovering that a serious labelling error had been made on the DNA sent for retesting to the defence experts, he responded that this fell outside the scope of quality-assurance monitoring!

An adversarial legal system is unhelpful for a developing area of forensic science. In contrast to the usual process of scientific advance, there is a tendency for those

involved to become locked into fixed positions because of previous testimony. Well-founded fears that openness about potential problems with the technique will be blown out of proportion by defence lawyers have resulted in something of a siege mentality among some forensic scientists. For example, several months before the Deen appeal, the prosecution's experts decided to retest the relevant DNA with a different probe. It transpired that the membrane from the original trial containing the defendant's DNA was missing. The crime-sample membrane was still available, as was another membrane containing the defendant's DNA. Not only was the defence not alerted to, nor invited to oversee, the additional testing, but subsequent reports made no mention of the loss and replacement of the original membrane. Deen's counsel learnt of this only immediately before the appeal. The resulting fiasco forced an adjournment. In addition, secrecy about a potentially unimportant substitution served to create the impression of a material concealment.

In some cases, subjective judgements are still an integral part of the profiling process. Those assessing DNA evidence should be aware of any such subjectivity, and of possible consequences. The DNA evidence in the Deen case was based on multi-locus, rather than the more common single-locus, probes. British forensic scientists had scored 10 matching bands, as a consequence of which they calculated a 'conservative' match probability of 1 in 700,000. At the appeal, a professor of forensic medicine from Germany testified that in his opinion there were 6 matching bands, 4 bands whose effect was neutral, and 2 discrepant bands. His assessment of the match probability was that it was no smaller than 1 in 33, and that the DNA evidence may even be exculpatory.

It is clear that some of the early claims for 'DNA fingerprinting' were unwarranted. Whether they were due to ignorance, optimism or commercial interests is less clear. Valid concern remains about the presentation and interpretation of DNA evidence. The costs of previous misunderstandings in unsafe convictions are impossible to assess, as are the undoubted benefits in terms of additional correct convictions. □

David J. Balding and Peter Donnelly are at the School of Mathematical Sciences, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

1. Lander, E. *Nature* **339**, 501–505 (1989).
2. Lewontin, R. C. & Hartl, D. L. *Science* **254**, 1745 (1991).
3. Roberts, L. *Science* **254**, 1721–1723 (1991).
4. Anderson, C. *Nature* **355**, 753–754 (1992).
5. Roberts, L. *Chance* **5**, 60–66 (1992).
6. National Research Council *DNA Technology in Forensic Science* (National Academy, Washington, DC, 1992).
7. MacLwain, C. *Nature* **367**, 101 (1994).
8. Koehler, J. J. *Jurimetrics* **34**, 21–39 (1993).
9. Anderson, C. *Nature* **359**, 349 (1992).
10. Thompson, W. C. *Judicature* **77**, 57 (1993).

More models of muscle movement

Physical models of phenomena such as muscle movement continue to be elaborated, but their relevance to the real world will be clearer only when the structure of the molecules concerned is better known.

It may be that last year's neatest measurement was that of the movement of kinesin molecules along the surfaces of microtubules (K. Svoboda, C.F. Schmidt, B.J. Schnapp & S.M. Block *Nature* **365**, 721; 1993). Svoboda and his colleagues from the late Edwin Land's Rowland Institute of Science in Cambridge, Massachusetts, showed, by direct optical measurement, that kinesin (like myosin, one of nature's agents for effecting mechanical motion) will, in the presence of ATP, move along a microtubule to which it is attached in consecutive steps about 8.2 ($\pm$ 1.1) nanometres long, often (in the real cell) dragging a cellular organelle behind it.

The neatness of the measurement, or at least its sophistication, is not questioned. Kinesin molecules were passively anchored to spherical silica beads 0.6 μ m in diameter, which were then loaded on to microtubules. The essence of the arrangement is a microscope for focusing a split laser beam onto two overlapping spots separated by 250 nm in such a way that the position of the sphere is at once made visible and constrained, or "trapped", within a region of a few hundred nanometres across. Svoboda and colleagues estimate that the restoring force (towards the centre of the overlapping laser spots) on a silica sphere near the edge of the field amounts to 5 pico-newtons, about enough to lift 0.005 micrograms against gravity.

The measurement is important as well as neat, as Jonathon Howard remarked at the time (*Nature* **365**, 696; 1993). For one thing, the average speed of movement turns out to be a sensitive function of the concentration of ATP; the more ATP, the more jumps a second, from a few to nearly a hundred. Second, and more immediately telling, the estimated size of each step along a microtubule is almost exactly that expected from what is known about the structure of tubulin, the essential ingredient of microtubules, which consists of two similar but distinct proteins anchored together head to tail, and which span just 8 nm.

Polymers of these dimers then associate laterally, perhaps 13 of them at a time, to form mechanically relatively rigid microtubules. The inference is that there is one binding site on each tubulin element. From this it is also possible to infer that each step taken along a microtubule requires the hydrolysis of one molecule of ATP.

As luck will have it, there is not yet much to say about the structure of the active heads of the kinesin molecules except that they are two-headed. But since Svoboda *et al.* appeared, the mechanisms by which myosin

molecules effect the relative movement of muscle fibres has been described even more graphically than can be done for kinesin by a further refinement of the same technique (J.T. Finer, R.M. Simmons & J.A. Spudis *Nature* **368**, 113; 1994). The hydrolysis of ATP induces a conformational change in the head which then induces movement by mechanical leverage. It is not unreasonable to suppose that something much like this happens with kinesin. The protein attachment of kinesin to a microtubule must change shape when it hydrolyses one molecule of ATP, and the net effect must be that it moves one 8 nm step along the microtubule.

But how does the kinesin "know" (anthropomorphically speaking) in which direction it should move? Quite separately from biologists' interests in what cells are up to, there has grown up a minor industry with a bearing on this question. Six months ago, Marcelo Magnasco from the Nippon Electric Company's research laboratory at Princeton and the Rockefeller University in New York described one way in which a linear motor, as in a muscle, might function (*Nature* **365**, 203; 1993), generating the energy needed to sustain a mechanical force "for free", as he put it, from its thermally fluctuating surroundings.

The guiding principle is that of an asymmetrical energy-well which is replicated indefinitely in some direction. It may, for example, be a saw-tooth structure in which the two edges of the teeth make different angles with the vertical. This, then, is a ratchet. Suppose then that there is a particle in one of the potential wells, that the particle itself is in communication with a heat bath at some fixed temperature (and so subject to random energy fluctuations) and that it is pulled in one direction or the other by a fixed force, not by itself sufficient to pull the particle in its own direction.

Magnasco sought to define the conditions in which the particle would move consistently in the opposite direction, against the force, doing work in the process. His conclusion was that if the ratchet is correctly oriented (with the more gently rising slope in the direction of the intended motion) and if the energy fluctuations are not strictly thermal (but "coloured", as they say in the trade), the particle would indeed move against the direction of the external force, doing work in the process, or acting as if it were a mechanical rectifier.

Now R. Dean Astumian and Martin Bier from the University of Chicago have taken the argument a step further (*Phys. Rev. Lett.*

72, 1766–1769; 14 March 1994). In their view (expressed after the appearance of Svoboda *et al.*), fluctuations of the forces to which the particle is exposed and fluctuations of the height of the potential barrier may be more realistic than other assumptions. The second model is supported by some speculation about the way in which electrical charges carried by ATP may neutralize the natural charge pattern of the protein substrate (in the case of kinesin, the tubulin polymer filaments).

Certainly the numerical agreement between the outcome of these calculations and the measurements of Svoboda is remarkable. Astumian and Bier also conclude, for example, that at the highest rates of motion of kinesin along a microtubule filament, more than one molecule of ATP may be needed to accomplish each step. On the other hand, the prediction that lower speed might make still more economical use of ATP appears not to be borne out by Svoboda *et al.*'s observation of the uniform motion of kinesin molecules at low ATP concentrations.

What matters in all this is the degree to which the models are likely to be useful not merely in accommodating the data now certain to be harvested in large quantities, but in suggesting whether some observations are likely to be more pertinent than others. That polymerized tubulin filaments must have the structure of a ratchet seems to go without saying, and accords with the idea that the tubulin dimer consists of two similar proteins joined head-to-tail. That is why, no doubt, it turns out that intact microtubules turn out to be unidirectional.

The obvious gap in the argument is the glaring uncertainty about the function of ATP in real life. If microtubules are ratchets and if the role of ATP hydrolysis is, say, to bring the two heads of kinesin together, it is easy to believe that only the relaxation step (following hydrolysis) will be controlled by the ambient fluctuations, thermal or otherwise. Even so, those embarking on the intricate optical measurement of contractile molecules such as those pioneered by Svoboda *et al.* could do worse than read the modest literature of the model-builders.

Meanwhile, the objection that the model-builders are pushing violations of the second law of thermodynamics should be suppressed. "Coloured" noise is itself a sign of a system out of equilibrium. The interesting question is rather that of how the hydrolysis of ATP by contractile proteins can simulate the appearance of fluctuations that are not boringly Gaussian.

John Maddox

Pan-editing in the beginning

Michael W. Gray

TRYPANOSOMATID protozoa engage in some odd biochemistry, none odder than RNA editing. In the mitochondria of these organisms (*Crithidia*, *Leishmania* and *Trypanosoma*, for example), protein-coding transcripts are altered by site-specific deletion of certain genome-encoded uridine residues and addition of other, non-encoded Us (ref. 1). This process re-tailors an initial primary transcript ('pre-edited RNA') so that the final product specifies a complete, functional protein. Besides localized editing of a particular region of the pre-edited RNA (5'-editing), trypanosomatids carry out a spectacular type of 'pan-editing' that in some cases accounts for 50 per cent or more of the mature sequence of messenger RNA¹. On page 345 of this issue², Maslov *et al.* present evidence that, contrary to what one might expect, pan-editing is an ancient trait, rather than a recently acquired one, within the trypanosomatid lineage.

The evolution of RNA editing systems — how they emerged and why they persist — is a challenging problem in evolutionary biology³. To date, U addition/deletion editing has only been found in mitochondria, and only within the order Kinetoplastida (which comprises the trypanosomatids and the bodonids and related cryptobiids). However, other types of editing of mitochondrial mRNA have been documented⁴, as has editing of mitochondrial transfer RNAs⁵⁻⁷ and ribosomal RNA⁸. The variety and mechanistic distinctiveness of these systems, coupled with their highly restricted occurrence, strongly suggests that most if not all of them are traits recently derived in evolution³.

Distribution

But how recently? In their study, Maslov *et al.* examine the distribution of pan-editing and 5'-editing in a phylogenetically disparate group of trypanosomatids. One might suppose that if U addition/deletion editing is a derived trait within this lineage, pan-editing would be restricted to certain of the more recently evolved genera. If the machinery for RNA editing arises before editable sites in RNA actually appear³, then the original genes would have been intact, and would only have begun to degenerate in sequence as their transcripts were able to be 'corrected' at the RNA level. So one might anticipate that any U addition/deletion editing would initially have been quite limited, both in terms of the set of transcripts altered and the number of sites affected within those transcripts. The system would, however, have been able to

expand in scope as the editing process became fixed and began to evolve further. It follows from such reasoning that deeply diverging branches within the trypanosomatids should not show evidence of pan-editing, which the above scheme predicts would have arisen in later diverging branches.

In fact, what Maslov *et al.* find is precisely the opposite. The authors first derive nuclear 18S rRNA trees of the trypanosomatid lineage, obtaining essentially the same topology using several different treeing algorithms. This result, in concert with high bootstrap re-sampling values for various branches in the tree, testifies to the robustness of their phylogeny. Next they examine the state of editing among their trypanosomatid subjects for three mitochondrial genes for which both pan-editing and 5'-editing were previously reported. They find that pan-edited forms of these genes are mainly found in the early diverging branches of the phylogenetic tree, whereas less highly edited forms make their appearance among the more recently evolved species. (Curiously, although Maslov *et al.* tell us that RNA editing occurs in the single cryptobiid — and most deeply diverging kinetoplastid — included in their tree, they do not tell us which kind it is: 5'-editing or pan-editing.)

These findings indicate rather clearly that pan-editing is the ancestral state of U addition/deletion editing within the trypanosomatid protozoa, leading the authors to suggest that "editing may be a more primitive mechanism than previously thought". Indeed, these observations might give heart to those who would like to believe that RNA editing is a retained relic of an ancient RNA world. A mechanistic parallel between RNA self-splicing and U addition/deletion editing, both of which occur by way of a concerted, two-step transesterification mechanism, has been recognized^{9,10}, and a yeast mitochondrial self-splicing intron is able to carry out a series of reactions *in vitro* that closely resemble the guide RNA-based model of RNA editing in trypanosomatid mitochondria¹¹. On the face of it, these similarities seem to strengthen the view that RNA splicing and RNA editing may be evolutionarily related, and that both may trace their origin to a pre-biotic world.

On the other hand, one should not be tempted to make too great an evolutionary leap backwards at the present time. It is a long way back to the prokaryotic endosymbiont that was the likely ancestor of the mitochondrion of trypanosomatids and other eukaryotes¹², and it is a much

further extrapolation still to the last common ancestor of the contemporary biological world, and from there to a primordial (and still largely hypothetical) RNA world. At the moment, there is no indication that U addition/deletion editing extends beyond the trypanosomatid mitochondrial lineage, let alone to the α -Proteobacteria, the closest bacterial relatives of mitochondria¹². A particularly pertinent question is whether this type of RNA editing occurs in the mitochondria of the nearest evolutionary neighbours of the trypanosomatid protozoa, the euglenoid flagellates. If it does, it would provide critical support for a truly ancient origin of U addition/deletion editing.

Relationship

A final issue addressed by Maslov and co-workers is the relationship between pan-editing and 5'-editing. The authors conclude that, at several times in the course of trypanosomatid evolution, pan-edited genes were replaced by 5'-edited ones. A plausible explanation is that complementary DNA copies of partially edited mRNAs were incorporated into the mitochondrial genome, displacing the original pan-edited 'cryptogenes'. Such a mechanism has previously been invoked to account for marked differences in the extent of C-to-U substitution editing of homologous mitochondrial transcripts in different angiosperm species⁴. What we may be witnessing in these cases is a sort of evolutionary tug-of-war between RNA editing and reverse transcription: the first permitting the progressive radicalization of mitochondrial gene sequences, the second mediating the conservative restoration of the same sequences. It seems that in their quest for biological novelty, mitochondria may be allowed to go only so far. □

Michael W. Gray is in the Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.

- Hajduk, S.L., Harris, M.E. & Pollard, V.W. *FASEB J.* **7**, 54–63 (1993).
- Maslov, D.A., Avila, H.A., Lake, J.A. & Simpson, L. *Nature* **368**, 345–348 (1994).
- Covello, P.S. & Gray, M.W. *Trends Genet.* **9**, 265–268 (1993).
- Gray, M.W. & Covello, P.S. *FASEB J.* **7**, 64–71 (1993).
- Loneragan, K.M. & Gray, M.W. *Science* **259**, 812–816 (1993).
- Janke, A. & Pääbo, S. *Nucleic Acids Res.* **21**, 1523–1525 (1993).
- Maréchal-Drouard, L., Ramamonjisoa, D., Cossett, A., Weil, J.H. & Dietrich, A. *Nucleic Acids Res.* **21**, 4909–4914 (1993).
- Mahendran, R., Spottiswood, M.S., Ghate, A., Ling, M.-I., Jeng, K. & Miller, D.L. *EMBO J.* **13**, 232–240 (1994).
- Cech, T.R. *Cell* **64**, 667–669 (1991).
- Blum, B., Sturm, N.R., Simpson, A.M. & Simpson, L. *Cell* **65**, 543–550 (1991).
- Mueller, M.W., Hetzer, M. & Schweyen, R.J. *Science* **261**, 1035–1038 (1993).
- Gray, M.W. *Curr. Opin. Genet. Devel.* **3**, 884–890 (1993).

The chemistry of crystalline sponges

John Meurig Thomas

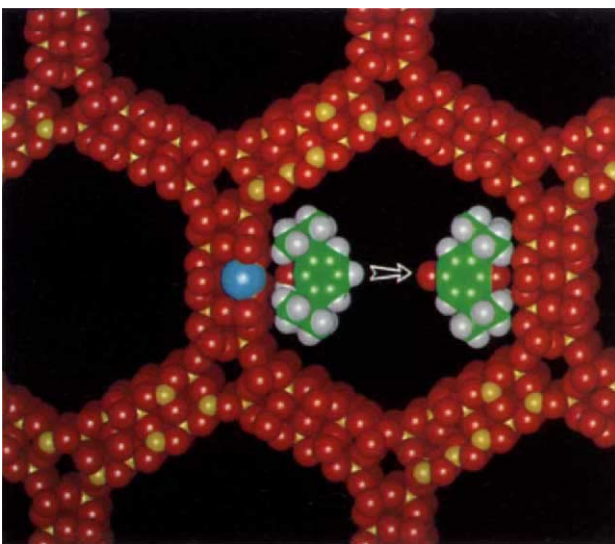
THE prospect of producing open-structure networks of a wide variety of inorganic materials, with apertures in the range 20 to 200 Å diameter, is brought a stage nearer by the work of Galen Stucky, Ferdi Schüth and their co-workers, reported on page 317 of this issue¹. And Tanev *et al.*, also in this issue, report the synthesis of a new titanium-containing mesoporous sieve². Such mesoporous ordered solids are likely to be of great practical value in a host of applications in the physical and biological sciences, in engineering and conceivably in medicine. Their mode of formation also sheds new light on the important phenomenon of bio-mineralization.

Many of the applications will be logical extensions of the myriad uses to which microporous, open-structure solids of aperture dimensions in the range 4 to 15 Å are already put. Microporous, micro-crystalline zeolites, as well as ALPOs (aluminium phosphates with molecular-sieve, open-framework structures), MeALPOs (where Me is a divalent metal ion such as Mg, Zn or Co that replaces some of the Al), so-called SAPOs (where Si replaces some of the P) and the KTP family (potassium titanyl phosphate), are used extensively in separation science, in heterogeneous catalysis, in nonlinear optics and electronic devices^{3,4}, and in 'ship-in-bottle' syntheses⁵. Much larger molecular entities, like the bulky species present in viscous oil, will be more readily processed, separated and converted by meso- rather than by microporous absorbents and catalysts into the interstices of which large molecules cannot penetrate. But very many other outlets seem likely for the mesoporous, ordered materials which are the result of Stucky and co-workers' generalized strategy for their preparation.

Mesoporous solids are particularly attractive as uniquely appropriate supports for immobilized enzymes. Nearly 80 years ago the American chemists Nelson and Griffin⁶ used active charcoal — which has little or no ordered structure and, in retrospect, a disadvantageously wide pore-size distribution — as the inert support for the enzyme invertase which converts cane sugar into glucose and fructose. This was the prelude to the later imaginative work of Katchalski-Katzir⁷, who devised several superior methods of immobilizing enzymes without adversely affecting their catalytic apparatus or performance. Thanks to such work immobilized enzymes are nowadays used^{8–10} extensively for the industrial production of high-

fructose corn syrup (from immobilized glucose isomerase) and 6-aminopenicillanic acid (from penicillin amidase).

Given that the inner linings of the cavities and channels in Stucky and Schüth's new inorganic mesoporous solids are replete with pendant hydroxyl groups (as described below), it should^{9,10} be possible to anchor at their interior surfaces appropriate enzymes that will be

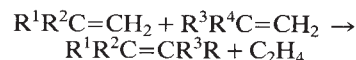


Computer graphic representation of a titanium-containing hexagonal mesoporous silica catalyst which selectively oxidizes 2,6-di-tert-butyl phenol into the corresponding quinone in the presence of hydrogen peroxide (see Tanev *et al.*², page 321 of this issue). Colour code: red, oxygen; yellow, silicon; blue, titanium; white, hydrogen; green, carbon. (R. G. Bell and J. M. Thomas, Davy Faraday Research Laboratory, Royal Institution.)

situated in precisely defined chemical environments rather different from those present in the formerly favoured porous diatomaceous (kieselguhr) earth, one of the other inorganic supports for immobilized enzymes.

'Crystalline sponges' of the appropriately chosen inorganic material will be attractive in other catalytic contexts. It has already been demonstrated that ordered

mesoporous tungsten oxide may be prepared; and surfaces of WO₃ are known to be effective catalytically in the metathesis of alkenes



This is an important reaction in the petrochemical industry in that, *inter alia*, it enables bulky alkenes to be readily prepared from less bulky ones. And with the pendant OH groups (in a mesoporous solid of composition close to WO_{3-x}(OH)_{2x}) it is not difficult to engineer means of anchoring other redox elements (rhodium for example¹¹) that could confer

bifunctional catalytic activity, such as the added facility to hydrogenate or dehydrogenate organic reactants. 'Tea-bag' catalysts¹², where the incarcerated catalyst is not actually anchored to the inner walls but rattles around within the cages, may also become a reality with these new ordered mesoporous solids.

Varieties of open-structure microporous (molecular sieve) solids based on synthetic and naturally occurring zeolites (aluminosilicates), ALPOs, MeALPOs and SAPOs continue to grow both in regard to structural type and in the range of elements of the Periodic Table that may be incorporated into the frameworks. Methods of synthesis of these materials — the product of ingenious adaptation of classical methods devised by Barrer¹³ and Breck¹⁴ — have succeeded in nearly doubling the cavity dimensions of the first generation of synthetic zeolites over a period of nearly 40 years. And exploitation^{15–18} of Barrer's original method of pillaring layered aluminosilicate minerals like montmorillonite and beidellite has yielded inorganic solids

- Huo, Q. *et al.* *Nature* **368**, 317–321 (1994).
- Tanev, P. T., Chibwe, M. & Pinnavaia, T. J. *Nature* **368**, 321–323 (1994).
- Cox, S. D., Eier, T. E. & Stucky, G. D. *Chem. Mater.* **2**, 609 (1990).
- Terasaki, O. *et al.* *Nature* **330**, 58–60 (1987).
- Lem, G. *et al.* *J. Am. chem. Soc.* **115**, 7009 (1993).
- Nelson, J. M. & Griffin, E. G. *J. Am. chem. Soc.* **38**, 1109 (1916).
- Katchalski-Katzir, E. *Trends Biotechnol.* **11**, 471 (1993).
- Tanaka, A., Tosa, T. & Kobayashi, T. (eds) *Industrial Applications of Immobilized Enzymes* (Dekker, New York, 1993).
- Lilly, M. D. *Chem. Engng Sci.* **49**, 151 (1994).
- Wisdom, R. A., Dunnill, P., Lilly, M. D. & Macral, A. *Enzyme Microb. Technol.* **6**, 443 (1984).
- Schwartz, J. *Accts chem. Res.* **18**, 302 (1985).
- Thomas, J. M. *Phil. Trans. R. Soc. A* **333**, 173 (1990).
- Barrer, R. M. *Zeolites and Clay Minerals* (Academic, London, 1978).

- Breck, D. W. *Zeolite Molecular Sieves: Structure, Chemistry and Use* (Wiley, New York, 1974).
- Vaughan, D. E. W. in *Zeolites: Facts, Figures, Future* (eds Jacobs, P. A. & van Santen, R. A.) 95 (Elsevier, Amsterdam, 1989).
- Diddams, P. A. *et al.* *J. chem. Soc. chem. Comm.* 1340 (1984).
- Thomas, J. M. in *Intercalation Chemistry* (eds Jacobsen, A. J. & Whittingham, M. S.) 56 (Academic, New York, 1982).
- Pinnavaia, T. J. *Science* **220**, 365 (1983).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartulli, J. C. & Beck, J. S. *Nature* **359**, 710 (1992).
- Beck, J. S. *et al.* *J. Am. chem. Soc.* **114**, 10834 (1992).
- Corma, A., Navarro, M. T. & Perez Pariente, J. J. *chem. Soc. chem. Comm.* 147 (1994).
- Notari, B. in *Innovation in Zeolite Materials Science* (eds Grobet, P. J., Martins, J. A. & Jacobs, P. A.) 143 (Elsevier, Amsterdam, 1987).
- Bonfield, W. *Proc. R. Inst. Great Britain* **61**, 159 (1989).

with effective (but not sharply defined) pore apertures of some 20 Å or so.

A great advance came nearly two years ago, when Kresge *et al.*¹⁹ showed how to prepare ordered, mesoporous aluminosilicate molecular sieves "using a liquid-crystal template mechanism". The result was regular arrays of uniform channels (see Fig. 1 of ref. 1), the dimensions of which could be tailored (in the range 16 Å to 100 Å or more) by adroit "choice of surfactant, auxiliary chemicals and reaction conditions". The mesoporous molecular sieve was produced by hydrothermal reaction of aluminosilicate gel in the presence of a quaternary ammonium surfactant, followed by calcination to drive away the contained template (some 40 per cent of the original surfactant). The surface area of such a crystalline sponge was in excess of 1,000 m² g⁻¹.

Although the reliability of Kresge and colleagues' preparation method for mesoporous silicas has since been frequently vindicated, their seductively simple model of the liquid-crystal templating mechanism has not stood up to rigorous experimental tests. Stucky and colleagues are in no doubt that it is unnecessary to form liquid-crystal micellar phases (portrayed as hexagonal arrays in Fig. 5 of ref. 19). They worked¹ with low concentrations such that no liquid crystal phase could form, and nonetheless generated the ordered mesoporous structures analogous to those described above^{19,20}.

It turns out that charge-density matching of the interlocking interfaces of the organic (surfactant) and inorganic regions in the nutrient liquor from which the novel structures emerge is all-important. The strategy unfolded in the paper by Stucky, Schüth and co-workers¹ makes it possible to prepare ordered mesoporous structures of the oxides of iron, antimony, lead and zinc as well as of the silica hexagonal mesoporous solids hitherto reported^{19,20}.

The feasibility of synthesizing mesoporous silicas with some redox element such as titanium incorporated into their inner surfaces is an idea that has occurred to several groups^{2,21}. After all, the remarkable properties of titanium silicalite, in which Ti(IV) ions are inserted into microporous silica thereby converting it into a highly selective oxidation catalyst for the generation of hydroquinone from phenol, is already²² harnessed industrially by the Enichem Company in Italy, where such work first began. Corma and his colleagues²¹, using their Ti-incorporated mesoporous silica (pore diameter roughly 20 Å) have selectively oxidized rather large organic molecules such as norbornene.

And Tanev *et al.*, elsewhere in this issue², report the synthesis of a mesoporous silica molecular sieve with site-isolated titanium. This 'sieve catalyst' is of such dimension that it can selectively

oxidize the bulky 2,6-di-tert-butyl phenol (shown inside a catalytic cavity of 20 Å diameter in the figure on the previous page) into the corresponding quinone.

Numerous other highly selective organic conversions can obviously be effected by these new mesoporous catalysts. But there are many noncatalytic applications, including the production of nanocomposite electrochromics or solid electrolytes, to which we may look forward. And given

the varied nature of the materials (including metal phosphates) which can now be fashioned into exceptionally porous states, it is not idle to speculate that these new materials may prove viable as biological implants²³. □

Sir John Meurig Thomas is at the Royal Institution, 21 Albemarle Street, London W1X 4BS and at Peterhouse, University of Cambridge, Cambridge CB2 1QY, UK.

GEOCHEMISTRY

A noble endeavour

Frank A. Podosek

WHAT common link is there between the life history of meteorites, the subtle composition of the Earth's atmosphere and a rare form of nuclear decay? The answer, as participants at a recent meeting* ably demonstrated, is the insight that noble gas research can bring to all of these areas. And a notable theme at the meeting was the continuing 'mainstreaming' of a discipline that many nonspecialists have considered inaccessible and arcane.

One problem of 30 years' vintage is the nature of the anomalous meteoritic component known as Xe-HL because of its strikingly high abundances, relative to normal xenon composition, of both the heaviest and the lightest Xe isotopes. Xe-HL is now viewed as having been carried into the Solar System in interstellar diamonds, still preserved in some primitive meteorites.

The roughly constant proportions of the H and L enrichments in prior observations have suggested that they are somehow coupled, and this has long been a puzzle. The H component must be produced in environments of high neutron density, and so is characterized by enrichment of neutron-rich isotopes, whereas the L component reflects an environment of photodisintegration or proton addition, which boosts the proportion of neutron-poor isotopes. It is thus unclear why these two components should apparently have come to reside in the same population of grains.

Now, new data (J. D. Gilmour and G. Turner, Manchester University) indicate that the H and L components are not actually so closely coupled. They are released in different proportions in stepwise heating, indicating that they are not trapped in the same mineral locations. Confirmation of these results may bring changes in the present view of pre-solar grains and isotopic components preserved in meteorites — for instance, what other isotopic anomalies may accompany those

in the noble gases, and what kinds of stars may have produced interstellar diamonds.

The instrument responsible for these results is itself a noteworthy development: a new mass spectrometer (dubbed RELAX) which achieves a reportedly hundredfold improvement in sensitivity for Xe. It does so through a combination of sample concentration by refrigeration, ionization by a laser tuned to a resonant wavelength for Xe, and time-of-flight analysis allowing collection of many different ion species per laser pulse — essential for these tiny samples.

Another issue which has vexed noble gas geochemists essentially since there were any is the attempt to relate noble gas compositions observed in meteorites and lunar samples to those of the Earth's atmosphere. Particularly for Xe, this has been a problem that has never been convincingly resolved.

One approach to the problem postulates the existence of an underlying primitive Xe component most commonly designated P- or U-Xe, the idea being that meteoritic Xe consists of U-Xe plus superposed components such as Xe-HL, whereas atmospheric Xe consists of strongly mass-fractionated U-Xe plus different superposed components such as Xe from spontaneous fission of actinides. But U-Xe is only hypothetical, its composition determined by extrapolation of isotopic correlations; many have doubted its existence as anything other than a mathematical construct.

At the meeting two reports (O. Eugster *et al.*, University of Bern; K. Nagao, Okayama University) asserted, on the basis of isotopic relationships, that a physical component at least very like U-Xe is actually present in some differentiated meteorites. If U-Xe is real it could have a substantial impact on models for meteorite observations and especially models for evolution of the Earth's atmosphere. It would, for example, provide a quantitative basis for estimating just how much and what kind of actinide fission Xe may be mixed into the atmosphere.

*International Workshop on Noble Gas Geochemistry and Cosmochemistry (Yamada Conference XXXVIII), Kyoto, Japan, 17–19 January 1994.

Jewish lysosomes

Jared M. Diamond

Progress has also been made on one of the oldest problems confronted by noble gas techniques: double-beta ($\beta\beta$) decay, a phenomenon of great interest in theoretical physics because of the limits it can set on the basic properties of the neutrino and on electroweak interactions. $\beta\beta$ decay is very slow, and few positive cases are known either from direct laboratory observation or through the geochemical method of daughter isotope accumulation (the latter is possible only when the daughter isotope is exceedingly scarce, in other words a noble gas).

The first evidence that $\beta\beta$ decay occurs at all (excess ^{130}Xe from decay of ^{130}Te in Te ores) came from noble gas studies 45 years ago, and a handful of other nuclides have since been found to undergo $\beta\beta$ decay. But some controversies have persisted for decades. One of these is whether ^{128}Te also decays. New data (F. A. Podosek *et al.*, Washington University) for ^{128}Xe excesses in Te ores indicate that it indeed does.

Another controversy still persists. Whereas we would argue for a 'best-estimate' ^{130}Te half-life of 2.7×10^{21} years, it can still be argued (as by N. Takaoka, Kyushu University) that the half-life is threefold lower (0.8×10^{21} years). The different half-lives lead to different constraints on neutrino properties, which follow from the absolute decay rate of ^{128}Te , in turn known by its ratio to the decay rate of ^{130}Te . As excellent noble gas data lead to a wide range of estimated half-lives, the problem evidently lies in the complex geological history of telluride ores.

As noted at the beginning, mainstream geoscience's interest in noble gases is also gathering momentum. One need only look at the work on developing noble gases as tracers for fluid flow in sedimentary basins, the accelerating pace of studies of noble gas partitioning and diffusion in silicates, and the relevance of noble gases in models of how the main terrestrial reservoirs could have evolved.

To pick out just one significant development, there is evidently growing acceptance of the notion that neon in at least the upper mantle is a 'solar' component (M. Honda *et al.*, Australian National University). The implications of the Earth's having solar-type noble gases in the first place are just beginning to be explored, but the difference between solar and atmospheric Ne ($^{20}\text{Ne}/^{22}\text{Ne}$ is about 25 per cent lower in air) is already influencing models for the generation of the atmosphere. □

Frank A. Podosek is in the Department of Earth and Planetary Sciences and in the McDonnell Center for the Space Sciences, Washington University, St Louis, Missouri 63130, USA.

A LONG-standing puzzle of human population genetics has been the incidence of a dozen autosomal recessive diseases at high frequencies in eastern European (Ashkenazi) Jews^{1,2}. Best-known are Tay-Sachs disease (which is always fatal) and Gaucher disease (which is often serious), both with heterozygote frequencies of several per cent. Why should Ashkenazi Jews, of all people, have been singled out by these genetic curses? Studies of Gaucher disease by Ernest Beutler and others continue to narrow the search for an evolutionary answer³⁻⁸.

One simple yet plausible explanation would be that Jews were singled out merely by chance. Deleterious genes may transiently rise to high frequencies in small populations through the founder effect or genetic drift. The history of eastern European Jewry, which began around the thirteenth century AD with the arrival of a modest number of immigrants from western Europe, would have lent itself to such an interpretation¹. Yet it has always seemed suspicious that three of the common Ashkenazi diseases (Tay-Sachs, Gaucher and Niemann-Pick) all involve lipid storage resulting from defects in three separate lysosomal enzymes (hexosaminidase A, glucocerebrosidase and sphingomyelinase respectively). That seems to suggest that lightning has struck Jewish lysosomes too often to be a coincidence. But the argument is not conclusive, because three similar defects might have been encountered by chance in at least one human population if enough populations were scrutinized for enough diseases.

The case against coincidence would be strengthened if it could be shown that each of those lysosomal diseases in turn had several origins among Jews. In 1979, in the book *Genetic Diseases among Ashkenazi Jews*, the geneticist Victor McKusick pointed out that "If there are two or more different alleles for the hex-A [Tay-Sachs] mutation, which achieved high frequency in the Ashkenazim, this finding would be a powerful argument for selection since chance would not be likely to raise two or more alleles to significant frequencies" (p. 313 of ref. 1).

In that same book several authors went on record as predicting that only a single allele would be found to account for the Ashkenazim's high frequency of Tay-Sachs or Gaucher disease. The first disease for which this prediction was proved wrong was Tay-Sachs, which is now known to be caused in Ashkenazi Jews by two distinct common mutations as well as by other rare ones⁹. New studies have now

reached a similar and even more striking conclusion for Gaucher, the commonest of all the storage diseases³⁻⁸. Five common mutations collectively account for about 97 per cent of Jewish Gaucher disease alleles. Each mutant allele has a measured or estimated heterozygote frequency of 0.05–2.8 per cent — far higher than could be sustained at equilibrium (in the face of deaths or debilitation from Gaucher) by recurrent mutations in the absence of any compensating selective advantage. Four of the mutant alleles occur in Jews in the context of only a single haplotype. That implies that the mutations arose in Jews recently, and only once, and have within a short time been pumped up to high frequencies by selection. Of those four mutations, all are rare or occur at much lower absolute frequencies in Jews than in non-Jews.

In effect, lightning has struck Jewish lysosomes not once, not three times, but at least eight times. If your house is hit by lightning while a neighbour's house goes unscathed, you may curse your bad luck, but if it happens eight times you should seek causative factors. In genetic terms, that means seeking some compensating advantage of the deleterious gene, most likely in the heterozygote, as illustrated by the familiar example of sickle-cell haemoglobin conferring resistance to malaria. Furthermore, at least one Jewish autosomal recessive disease that is not caused by a lysosomal enzyme defect — factor XI deficiency — has also turned out to involve several alleles each at high frequency¹⁰. What selective force could be acting on eastern European Jews but not on their non-Jewish neighbours? Specifically, what features of Jewish living conditions or lifestyle during the past 700 years could have conferred com-

1. Goodman, R. M. & Motulsky, A. G. (eds) *Genetic Diseases among Ashkenazi Jews* (Raven, New York, 1979).
2. Bonnè-Tamir, B. & Adam, A. *Genetic Diversity among Jews* (Oxford University Press, 1992).
3. Beutler, E. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5384–5390 (1993).
4. Beutler, E. *Am. J. Dis. Children* **147**, 1175–1183 (1993).
5. Tsuji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 2349–2352 (1988).
6. Beutler, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 10544–10547 (1991).
7. Beutler, E. *New Engl. J. Med.* **325**, 1354–1360 (1991).
8. Beutler, E. *Science* **256**, 794–799 (1992).
9. Neufeld, E. F. *J. biol. Chem.* **264**, 10927–10930 (1989).
10. Tuddenham, E. G. D. (ed.) *Haemostasis and Inherited Disease* 212–220 (Oxford University Press, 1994).
11. Myrianthopoulos, N. C. & Aronson, S. M. *Adv. exp. Med. Biol.* **19**, 561–570 (1972).
12. Petersen, G. M. *et al. Am. J. hum. Genet.* **35**, 1258–1269 (1983).
13. Spyropoulos, B. *et al. Am. J. hum. Genet.* **33**, 375–380 (1981).
14. Diamond, J. M. & Rotter, J. I. *Nature* **329**, 105–106 (1987).

pensatory value on the genes for Gaucher, Tay-Sachs, Niemann-Pick and factor XI deficiency?

This conundrum has stimulated much discussion in print^{1,2,11-16}, and even more discussion not in print. The present stage is still one of speculative search for testable hypotheses, and not yet one of definitive hypothesis-testing. For instance, could being crammed into tuberculosis-ridden urban ghettos have placed eastern European Jews under much stronger selection for resistance to tuberculosis than their predominantly rural Gentile neighbours? An unexpectedly low frequency of deaths from tuberculosis among Tay-Sachs heterozygotes has in fact been reported¹¹, but the evidence remains inconclusive¹³. A preliminary survey of Gaucher gene frequencies among Jewish tuberculosis patients failed to detect the hypothesized resistance to tuberculosis¹⁷.

A second hypothesis is selection in Jews for the intelligence putatively required to survive recurrent persecution, and also to make a living by commerce, because Jews were barred from the agricultural jobs available to the non-Jewish population¹. A related third possibility is sexual selection for increased reproductive success, because men with the qualities required to become rabbis were prized as husbands and would have tended to marry wealthy women capable of nourishing many children. Of course, these factors are hypothetical, and other explanations are possible.

Identification of the selective factors responsible is unlikely to be easy. Heterozygote frequencies for these Jewish recessive diseases are lower than peak frequencies for sickle-cell haemoglobin in Africa, hence the balancing heterozygote advantages required to sustain otherwise-deleterious genes at equilibrium are also slighter. The selective factors may be ones that have nearly vanished, such as bubonic plague, or ones that have decreased in impact, such as tuberculosis.

To overcome these difficulties, two possible research strategies, already attempted in Tay-Sachs studies^{11,12}, are to enquire about parental or grandparental geographical origins (within eastern Europe) of living bearers of Gaucher genes, and to enquire about the recollected causes of death of those parents or grandparents. No doubt, studies of natural selection and sexual selection in *Drosophila* will always be easier and less controversial. But selection in humans will remain more significant for understanding human disease and history — and more interesting. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA. This is his one-hundredth News and Views article.

Cosmology in a helium cell

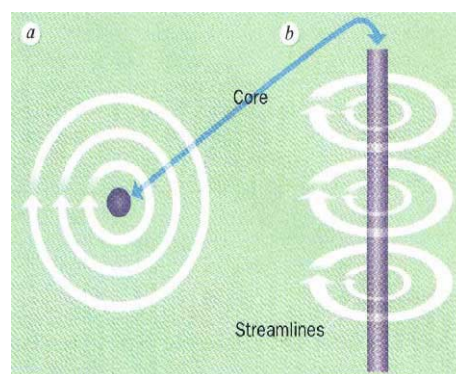
W. H. Zurek

LORD Rutherford's memorable phrase, "When a student in my laboratory mentions 'the Universe', I tell him it's time for him to leave", must have been ringing in the ears of Peter McClintock and his colleagues¹ as they worked on the experiment described on page 315 of this issue. They have defied the old 'Crocodile', as he was known to his colleagues, by demonstrating that vortex lines — tiny, rotating, quantized eddies of superfluid — are copiously created as a consequence of a rapid phase transformation from the normal to the superfluid phase of ⁴He. The experiment was suggested²⁻⁴ as an analogue test of the process of cosmic string creation proposed by T. W. B. Kibble⁵.

As the primaeval fireball expanded following the 'Big Bang', the Universe cooled, and this precipitated spontaneous symmetry-breaking phase transitions from the symmetric ('false') to the broken-symmetry ('true') vacuum. This process determined the nature of the fundamental interactions — the split from unified force into today's strong, electromagnetic, weak and gravitational forces. Spontaneous choice of the direction of magnetic field in a ferromagnet cooled below the critical temperature is a classic example of a phase transition. However, the field is more or less uniform in the domains of the resulting magnet because the transition occurs slowly, so that the choice of magnetization direction can be communicated throughout the sample. Moreover, even if the direction was initially non-uniform, the magnetic field can always disentangle itself into a uniform configuration in which its lines of force are 'combed' more or less in the same direction.

In the early Universe such uniform combing of the elementary fields is less likely. Immediately after the phase transition, the field that determines the choice of the broken symmetry can adjust to be approximately uniform only in regions that are causally connected — that is, on scales which are at most of the order of the distance travelled by the light since the Big Bang. Many such regions are enclosed inside our present-day Universe. The direction of the fields within each of them must have been selected independently, and at random.

As the causal horizon grows and different regions begin to interact, the orientation of the field lines attempts to adjust so as to minimize the total energy. Absolute minimum would be reached only if the field lines throughout the entire Universe were combed in the same direction. For most of the symmetry-breaking schemes of high-energy physics, such a state cannot be reached starting from an arbitrary random initial configuration.



a, Topologically stable configuration of the field lines in a plane. In the broken-symmetry phase lines point everywhere in a specific direction. When arranged as shown above, they trap some of the symmetric phase ('false vacuum') in the core of the configuration. Continuity demands that the core of such a defect should remain 'undecided' about how to break symmetry. *b*, Three-dimensional defect with the topology of a string. In superfluid helium this would be the flow pattern around a vortex line, with the normal fluid in the core. *a* can be thought of as a section through such a vortex line.

To see why, imagine field lines on a plane combed so as to point everywhere in the clockwise direction (part *a* of the figure). The lines near the origin, at the centre of the clock, will not be able to decide how to break the symmetry — which way to point — and the original symmetric phase will be trapped here. This configuration cannot be disentangled for topological reasons: the field lines cannot be aligned everywhere on the plane by means of a local intervention, that is, without doing something to the whole configuration, including lines arbitrarily far from the centre of the clock.

Such impossible-to-remove configurations frozen into the broken-symmetry phase are known as topological defects. Their inevitable emergence in the course of cosmological phase transitions is both the curse and the hope of cosmologists. Copious production of monopoles — three-dimensional configurations with field lines pointing out like bristles in a hedgehog — is thought to be inevitable in every plausible unified model of elementary interactions. Moreover, the monopoles must have been very heavy and strongly interacting. Their predicted

abundance appeared to threaten Big Bang cosmology until it was realized that an inflationary epoch in the prehistory of the Universe could dilute them to make their contribution to the matter content negligible.

By contrast, Kibble's suggestion was that cosmic strings — strings made of trapped symmetric phase, which would connect the centres of stacks of planes, each with a whirl of field lines as discussed before (part *b* of the figure) — could act as seeds for galaxies⁵. Much effort has been devoted to the study of the 'cosmic string scenario' of galaxy formation⁶. Although cosmic strings were (and still are, especially after COBE measurements of the microwave background inhomogeneities) an exotic alternative to more mundane ways of seeding structure in the Universe, they have brought to the fore an exciting and very concrete problem of topological defects and of dynamical, non-equilibrium aspects of symmetry-breaking phase transitions in condensed matter.

In superfluid helium cooled through its critical temperature, the direction of the symmetry-breaking field is related to the phase of the quantum wave function of the Bose condensate, and it is the vectors of the velocity field that must be combed into alignment. The random choice of Bose condensate phases and the corresponding velocity field left after the phase transition cannot be easily disentangled. Where different regions meet, eddies of rotating superfluid are left. These are quantized 'holes' in the broken-symmetry phase filled in with symmetric (normal, non-superfluid) liquid helium⁴.

Nearly ten years ago, intrigued by the cosmological connection and excited by the prospect of generating rotation simply by inducing a rapid phase transition in the superfluid, I convinced Los Alamos experimentalist Jim Hoffer and his student, S. S. Shiah, to attempt the 'cosmological experiment' in bulk superfluid. The aim was to expand liquid ⁴He rapidly through the transition and detect the resulting vortices acoustically. Preliminary results were encouraging but inconclusive, and both Hoffer and John Wheatley, who directed low-temperature research in Los Alamos at the time, felt that the predicted superfluid effect was simply too important (cosmological motivations notwithstanding) to claim its existence on the basis of a poorly controlled experiment. And a more sophisticated experiment, such as McClintock and co-workers now report, proved difficult to put together, especially following John Wheatley's untimely death.

The first experimental confirmation therefore came from work carried out at room temperature. Bernard Yurke and colleagues bypassed many of the experimental difficulties posed by superfluids by using liquid crystals⁷. The rod-

shaped molecules of liquid crystals are randomly jumbled together at high temperature, but as the temperature is lowered, they tend to align, running into all of the topological problems discussed before. Not only can liquid crystals be studied at room temperature, but the string-like topological defects can be seen with a microscope, so that the details of the phase transition can be studied almost with the naked eye as it takes place^{7,8}.

Liquid crystal experiments do indeed confirm the basic predictions of the cosmological scenario, but they deal with systems that are significantly more 'messy' than superfluids. Moreover, the predicted generation of vorticity — that is, rotation — as a result of a phase transition seems more startling than the liquid crystal defects.

So McClintock and colleagues' experiment¹, which confirms that quantized vortices can be generated in abundance, is not only an exciting venture into 'experimental cosmology', but a careful and intriguing exploration of poorly known non-equilibrium aspects of the dynamics of symmetry breaking in the course of superfluid phase transitions. The initial density of vortex lines (10^{11} m^{-2} or more) is enormous. It is close to predictions based on the idea that the characteristic domain size 'frozen out' during the pressure quench is set by the critical slowing of the superfluid dynamics near the phase transition^{2,4}. Using an estimate of 30 ms for the quench timescale and adopting renormalization group scaling of the critical slowing, one obtains a predicted density of about $2.7 \times 10^{12} \text{ m}^{-2}$.

Similar experiments could in principle be carried out in superconductors, where the relevant effective field theory is more complicated ('local gauge' in contrast to the 'global gauge' applicable in ⁴He) and involves gauge (magnetic) field. Local gauge theories are actually a closer analogue of theories considered in the cosmological context.

There are experimental opportunities aplenty to be explored, starting with detailed understanding of the non-equilibrium aspects of the superfluid phase transition, but naturally extending to other phase transitions, as the example of liquid crystals already illustrates. The dynamics of symmetry breaking might shed new light on the underlying physics — as McClintock and colleagues believe

A torrefying tale

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Lead pollution is generally considered a modern scourge, the unwelcome byproduct of industrialization over the past 200 years. Not so, say Ingemar Renberg *et al.* on page 323 of this issue — analyses of Swedish lake sediments show that the problem has been around since Roman times or earlier. Above, a woodcut from the sixteenth-century work *De Re Metallica* by Georgius Agricola shows labourers 'torrefying' (roasting) lead ores in wood-fired ovens. L.M.

Mary Evans Picture Library

will be the case for ⁴He — in superconductors (including the high-*T_c* variety, some of which should respond to rapid pressure change in a manner similar to ⁴He) and superfluid ³He, which offers a bewildering zoo of topological defects⁹.

Among my favourites for future experiments would be a rapid phase transition of the kind described in this issue, but carried out in an annular container. There, the prediction is², rapid phase transition should result in a net flow of the superfluid along the circumference of the annulus at a detectable velocity. Even if the law of conservation of angular momentum does not seem to be threatened — the resulting flow can be regarded as an amplified brownian motion² — I find the counterintuitive appeal of generating rotation simply as a result of a phase transition very hard to resist. □

W. H. Zurek is in the Theoretical Astrophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

- Hendry, P. C., Lawson, H. S., Lee, R. A. M., McClintock, P. V. E. & Williams, C. H. D. *Nature* **368**, 315–317 (1994).
- Zurek, W. H. *Nature* **317**, 505–508 (1985).
- Kibble, T. W. B. *Nature* **317**, 472 (1985).
- Zurek, W. H. *Acta Phys. Pol.* **B24**, 1301–1311 (1993).
- Kibble, T. W. B. *J. Phys.* **A9**, 1387–1398 (1976).
- Vilenkin, A. & Shellard, E. P. S. *Cosmic Strings and Other Topological Defects* (Cambridge Univ. Press, 1993).
- Chuang, I., Durrer, R., Turok, N. & Yurke, B. *Science* **251**, 1336 (1991).
- Bowick, M. J., Chandar, L., Schiff, E. A. & Srivastava, A. M. *Science* **263**, 943–945 (1994).
- Salomaa, M. M. & Volovik, G. E. *Rev. mod. Phys.* **59**, 533–613 (1987).

Hands across the divide

Ian P. Trayer

ON page 306 of this issue¹, Xie *et al.* present an atomic-resolution structure of the regulatory domain of the head of scallop myosin, adding to the riches of structural information already bestowed upon the contractile community in the 1990s²⁻⁵. The structure gives an intriguing twist to the calcium-binding protein story, makes one more reserved about database sequence-searching, and opens up another problem in our comprehension of allosteric structural changes.

One of the features of the crystal structure of the head of the myosin molecule (subfragment 1 or S1), characterized last year⁴, is the long carboxy-terminal heavy chain α -helix around which the two light chains are clamped sequentially. The



FIG. 1 The three domains of the myosin molecular motor. The motor domain (red) contains the nucleotide-dependent actin-binding site(s) and the ATP hydrolysis site. The light chain/calmodulin-binding domain (blue) is characterized by up to six repeated IQ sequence motifs, each capable of binding calmodulin or light chains; this may act as the 'lever' amplifying the structural changes in the motor domain into mechanical work in all myosins, and it doubles as a regulatory domain in some. Finally, the cargo domain (green) specifies the function and dimerization potential of the individual myosin. (See ref. 11 for an overview of the myosin superfamily.)

essential light chain (ELC) winds around the amino-terminal portion of the helix, and the regulatory light chain (RLC) around the carboxy-terminal portion where the head joins the rod-like shaft of the molecules.

For skeletal myosin, it has been postulated that this light-chain domain (see Fig. 1) acts as the conformational coupling device, amplifying smaller changes in the main ATP and actin-binding motor domain to perform mechanical work. It is equally clear from molluscan myosins that information transfer is a two-way process. In these tissues, Ca^{2+} regulates the actomyosin ATPase (and hence contraction) by binding directly to the light chain, whereas in vertebrate striated muscles the switch occurs by Ca^{2+} binding to the troponin complex on the thin actin filaments. One can begin to see how a molecular mechanism based on the steric blocking model could act in these vertebrate systems by interfering with the multisite docking between actin and

myosin⁶. The molluscan system, however, poses more subtle questions about how changes in the myosin S1 structure itself must occur to account for this regulation.

It was in the early 1970s that Szent-Gyorgyi and colleagues⁷ discovered that Ca^{2+} regulated molluscan adductor muscles by binding directly to the myosin light chains. This turned out to be a useful test system for light-chain function, as the RLC can be easily and reversibly removed. Because loss of this light chain was accomplished by loss of the Ca^{2+} -dependence of the ATPase activity, which could be restored by replacing this light chain, it became known as the regulatory light chain. It was puzzling, however, that the isolated light chain did not bind Ca^{2+} . Later, protein-sequencing work showed that the myosin light chains belonged to the same class of Ca^{2+} -binding trigger proteins as, for example, calmodulin, troponin C and vitamin D-dependent Ca-binding protein. Each possesses two pairs of EF-hand motifs⁸ joined by a central helix. This structural motif (a helix-loop-helix) was the Ca^{2+} -binding domain, provided, according to sequence comparisons of the various members of the group, certain constraints in primary structure occurred (see Fig. 2). From these criteria, scallop myosin light chains should possess two (out of eight) competent metal-binding loops — domain 1 of the RLC and domain 3 of the ELC. We now know why neither of these is involved in the specific binding of the regulating Ca^{2+} , although the former does provide the non-specific Mg^{2+} and/or Ca^{2+} site when associated with the heavy chain; removal of this divalent cation is likely to destabilize interaction between the RLC and the heavy chain¹, allowing RLC dissociation.

The structure of the scallop regulatory domain presented by Xie *et al.* is essentially the same as that of the corresponding region of chicken S1 (ref. 4). It was prepared from the parent myosin by proteolysis, and no chemical modifications were needed to crystallize it (giving added confidence to the rest of the chicken S1 structure). The light chains are more highly resolved, showing critical side interactions, and only short segments from the amino and carboxy termini are absent from the diffraction pattern.

The most unexpected feature is the location of the trigger Ca^{2+} -binding site in

domain 1 of the ELC — this should not be so by the 'established' criteria (Fig. 2) and is only accomplished by the loop of the backbone of the EF-hand wrapping closely around the ensnared Ca^{2+} so that four of the seven oxygen-donating ligands are provided by the backbone carbonyls. Indeed, two Asp residues (19 and 22) provide liganding oxygens from both their side-chain and carbonyl functions. The deformation of this loop around the Ca^{2+} ion is stabilized by a hydrogen-bonded network involving domain III of the RLC and the heavy-chain helix. This limited interaction is the only contact region between the two light chains, and appears to give rise to a bend of 40° or so in the heavy-chain helix at this point. The same bend and light-chain contact region also occurs in the chicken skeletal S1 structure^{4,9} which is in a permanently 'switched on' state. Xie *et al.* speculate that this may be the trigger point for the allosteric consequences of the removal of Ca^{2+} ; in the scallop myosin this could disrupt the ELC-RLC contact, allowing the helix to transmit these changes to the base of the ATP-binding site (around residue 700) where large conformational changes are known to occur during ATP binding/hydrolysis^{4,5,10}.

The new structure answers some questions but also raises some fascinating prospects. We now know where the Ca^{2+} binds and can readily see how removal of the RLC destroys Ca^{2+} regulation. In smooth muscle and non-muscle myosin II molecules¹¹, the role of Ca^{2+} is to activate a protein kinase (through binding to calmodulin) which phosphorylates the amino-terminal region of the RLC. Most of this is missing from this structure¹ (and from the S1 structure⁴), but in both cases the polypeptide chain is heading tantalizingly in the direction of the light-chain contact area. Could a common structural mechanism for regulation be emerging, in this case with the phosphorylation state of the RLC regulating inter-subunit contact?

	Helix	Loop	Helix
		X Y Z -Y -X	-Z
a	...MKDS	DKNNDGRID	FDEF...
b	...FSLF	DKDGDGTIT	TKEL...
c	...FELF	DFWDGRDGA	VDAF...
		↑ ↑ ↑ ↑	

FIG. 2 Sequence comparisons of Ca^{2+} -binding EF-hand motifs. Segments of site 4 of troponin-C (a) and site 1 of calmodulin (b) are typical representatives of the consensus sequence of the EF-hand motif and are compared with the Ca^{2+} -binding sequence of the site 1 of the scallop ELC (c). In a and b the coordinating residues are marked (X, Y, Z, -Y, -X, -Z) and all of these residues donate oxygens from their side chains to coordinate the Ca^{2+} , except the position -Y where it is the backbone carbonyl. Both oxygens of the side chain at position -Z are involved in coordinating the metal ion. In c all the liganding groups are provided from within the 9-residue loop, 4 from the main chain carbonyls (red arrows) and 3 from the side chains (green arrows).

These myosin II molecules from smooth and non-muscle cells also exhibit an interesting property in the dephosphorylated state, in which the rod portion of the molecule folds up (the so-called 10S myosin) and makes contact with the neck region of the head, trapping the ATP in the active site¹². The contact between the rod and the neck could be at this RLC-ELC interface, also disrupting their interaction. Even the scallop myosin in the absence of Ca^{2+} will fold in this way¹³.

As Rayment *et al.*⁴ have pointed out, the way in which the light chains clamp on to the heavy-chain helix closely resembles that in which calmodulin binds to its target peptides^{14,15}. This is not surprising given the homology between the light chains and calmodulin, and the fact that more unconventional cellular myosins¹¹ contain calmodulin rather than light chains in their neck region (Fig. 1). A consensus sequence motif, the IQ motif, has been identified as the heavy-chain binding site for these subunits. The structural basis for the interaction between the light chains or calmodulin and this motif has been further clarified by Xie *et al.*, and the sequence motifs extended to include another bulky hydrophobic residue separated by 12 residues from the Ile (IQXXRGXX-XXRXY(W)). This now encompasses all known calmodulin-binding sites, but it remains a mystery as to why light chains bind to myosin II molecules and calmodulin to the rest.

There are now at least nine groups in the myosin molecular motor family¹¹ and the basic design of the molecular motor can be perceived (Fig. 1). The number of calmodulins binding to the regulatory domain can be anything from one to six, presumably each capable of binding four Ca^{2+} . If the scallop myosin poses problems in understanding the structural basis of the single Ca^{2+} -binding trigger event, it makes the mind boggle at the possibilities that exist with the more conventional members of this superfamily. □

Ian P. Trayer is in the School of Biochemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

Pumping iron in the Pacific

Mark L. Wells

LAST November, in the largest scientific manipulative experiment in oceanographic history, 480 kilograms of iron was added to the surface waters of the Pacific Ocean 500 km south of the Galapagos Islands. The objective was to test the contentious suggestion that iron limits phytoplankton production and biomass in regions of the world's oceans otherwise rich in plant nutrients. Scientists eagerly gathered at a meeting last month* to hear the preliminary findings of this experiment, known as 'IronEx1'.

The surface waters of the northeast Pacific, the equatorial Pacific and the Southern Ocean — over 10 per cent of the world's ocean — are replete in major nutrients such as nitrate and phosphate but support comparatively low biomass. The keen interest in understanding this enigma stems in part from suggestions that sustained increases in photosynthesis in these regions might in the past have removed enough carbon dioxide from the atmosphere to induce periods of glaciation¹.

The suggestion that iron limits ocean productivity in these areas is not new². But the concentrations of iron involved are tiny, and only recently have experimental methods improved enough to avoid metal contamination during sampling or analysis. John Martin of the Moss Landing Marine Laboratory helped to pioneer these 'clean' methods and convincingly demonstrated in bottle experiments that nanomolar iron enrichments of high-nitrate, low-chlorophyll waters did indeed increase phytoplankton growth and biomass^{3,4}. A criticism often levelled at this 'iron hypothesis', though, is that bottle incubations may not adequately reflect the natural ecosystem response.

John Martin therefore orchestrated the scientific and logistical planning for a large-scale iron enrichment experiment in the open ocean. Tragically, his untimely death prevented him from seeing the outcome, but his plan proceeded under the joint leadership of Ken Johnson and Kenneth Coale.

In mid-November 1993, the RV *Columbus Iselin* arrived south of the Galapagos Islands, so laden with equipment and personnel that it could hold only half its freshwater reserve. Iron and a highly sensitive inert tracer (SF_6 ; see ref. 5) were pumped into the propeller wash as the vessel steamed to and fro across an 8×8 km field over 24 hours, raising iron concentrations from the ambient (about 0.05 nM) to about 4 nM. The 'patch' was subsequently traced chemically by flow

injection analyses of iron and SF_6 in water from the ship's pumping system.

The initial response of the ecosystem response was mapped from the pumped water while the ship was under way and from hydrographic stations occupied each day both inside and outside the patch. Overhead, optical measurements from NASA's aeroplane-borne P-3 Orion laboratory looked for changes in phytoplankton pigments.

On the fifth day of the 9-day experiment, a low-salinity front moved into the region, subducting the patch to a depth of 30–35 m, where it remained confined to a 5–10-m-thick interval above the thermocline but still within the photic zone. By this time, iron concentrations within the patch had decreased from their initial values of about 4 nM to less than 0.2 nM (K. Johnson and K. Coale, Moss Landing Marine Laboratory).

And the response? From the preliminary findings, it seems that phytoplankton growth was indeed stimulated. Photosynthetic efficiency increased within 12 to 24 hours after the iron had been added, and reached maximal rates after 2 to 3 days (Z. Kobler and P. Falkowski, Brookhaven National Laboratory). By this point, primary production rates and chlorophyll concentrations had more than doubled (R. Barber, Duke University) and the patch was readily discernible by airborne laser-induced fluorescence measurements (F. Hoge, NASA Goddard Space Flight Center). Clearly, one of the central tenets of the iron hypothesis had been confirmed.

But the magnitude of the ecosystem response fell a long way short of that predicted from the experiments carried out in bottles. Chlorophyll concentrations increased to only 0.7 μg per litre (Barber) in contrast to the peak concentrations of about 6 μg per litre typically found in bottle incubations. Furthermore, the main growth appeared to be that of phycoerythrin-containing picophytoplankton less than 2.0 μm in size (Barber), whereas in bottle experiments the overwhelming response is from larger, chlorophyll-containing diatoms. These shipboard results have yet to be reconciled with contrasting airborne fluorescence measurements which showed that chlorophyll — rather than phycoerythrin, a different photosynthetic pigment — dominated within the patch (Hoge).

The increase in phytoplankton biomass caused ammonia concentrations to decrease by a factor of 3, but ambient nitrate concentrations fell only imperceptibly. In addition, total carbon dioxide decreased by only $\sim 6 \mu\text{M}$ and CO_2 fugacity by only

- Xie, X. *et al.* *Nature* **368**, 306–312 (1994).
- Kabsch, W. I., Mannherz, H. C., Suck, D., Pai, E. & Holmes, K. C. *Nature* **347**, 37–44 (1990).
- Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. *Nature* **347**, 44–49 (1990).
- Rayment, I. *et al.* *Science* **261**, 50–58 (1993).
- Rayment, I. *et al.* *Science* **261**, 58–65 (1993).
- Lehman, W. *et al.* *Nature* **368**, 65–67 (1994).
- Kendrick-Jones, J., Lehman, W. & Szent-Gyorgyi, A. G. *J. molec. Biol.* **54**, 313–326 (1970).
- Kretzinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248**, 3313–3326 (1973).
- Reedy, M. K. *Structure* **1**, 1–5 (1993).
- Yount, R. G. *J. Muscle Res. Cell Motil.* **14**, 547 (1993).
- Bement, W. M. & Mooseker, M. S. *Nature* **365**, 785–786 (1993).
- Cross, R. A., Jackson, A. P., Citti, S., Kendrick-Jones, J. & Bagshaw, C. R. *J. molec. Biol.* **203**, 173–181 (1988).
- Ankret, J., Rowe, A. J., Cross, R. A., Kendrick-Jones, J. & Bagshaw, C. R. *J. molec. Biol.* **217**, 323–335 (1991).
- Ikura, M. *et al.* *Science* **256**, 632–638 (1992).
- Meador, W. E., Means, A. R. & Quiocho, F. A. *Science* **257**, 1251–1255 (1992).

*Ocean Sciences Meeting, San Diego, California, 21–25 February 1994.

~10 μatm from their original values after 4 days (K. Van Scoy, University of East Anglia; F. Millero, University of Miami). The decreases are about a tenth of those expected if photosynthesis had used up all the available nitrate.

Not surprisingly, the debate continues on the causes for the difference between bottle-scale and mesoscale iron enrichments. Although we shall learn more when the final results become available, it probably will not be possible to solve this puzzle without repeating the experiment. Two loss terms were not well characterized: the loss of phytoplankton through zooplankton feeding within the patch, and the time-dependent loss of available iron in surface waters. In defence of the planners, however, no more equipment could have been brought on board without risk of capsizing the vessel.

The grazing question is a tricky one. Based on bottle experiments, the prevailing wisdom has been that stocks of the smaller picophytoplankton are limited by the tiny marine animals that 'graze' on them, and the larger phytoplankton are iron-limited. So the apparent increase in picophytoplankton within the patch came as a surprise.

Part of the explanation for the low response from larger phytoplankton may be increased grazing by zooplankton normally excluded or damaged in most bottle experiments. Microscopic examination of seawater samples during IronEx showed about a 50 per cent increase in microzooplankton (Barber), and net tows indicated that mesozooplankton which normally migrate daily to the ocean surface from deeper waters were remaining there to profit from the extra food supply (Coale) — at the meeting, Bill Sunda (National Marine Fisheries Service, NOAA) likened this behaviour to that of seagulls flocking in to feed.

So consumption by mesozooplankton could be reducing the concentration of phytoplankton below what is expected from bottle experiments. It may be that these larger zooplankton also feed on the smaller grazers and that the lifting of grazing pressure allows the picophytoplankton to capitalize on the higher iron availability.

Understanding the rate at which available iron is lost from the patch is crucial for explaining the extent of the ecosystem response. In contrast to what occurs in 'contained' experiments, total iron concentrations within the patch decreased by about 90 per cent by the fifth day of the open-ocean experiment (Johnson). Even then, we cannot quantify iron availability rigorously because current iron analysis techniques probably overestimate the biologically available fraction⁶. Repetitive fertilization will probably be necessary before direct comparisons to bottle incubations become meaningful. The

pulse of iron input appears to have been sufficient to 'kick-start' the ecosystem, only for it to stall as the iron ran out.

Although grazing effects and diminishing iron availability probably contributed to differences between the bottle-scale and mesoscale iron enrichments, several alternative hypotheses were raised at the meeting (F. Morel, Massachusetts Institute of Technology). These included indirect effects arising from reactive photochemical products created by iron photolysis, and adsorption of other metals onto the colloidal iron oxides formed upon iron enrichment, which then aggregated and sank from surface waters. Finally, bottle incubation results can differ from site to site (M. Wells, University of California, Santa Cruz), and the Galapagos region is several thousand kilometres from where most of the bottle experiments were conducted.

Has the 'iron hypothesis' been verified? So far, the answer is a qualified yes. Phytoplankton production and biomass did increase upon iron enrichment, although the results differed from those of bottle experiments. Nonetheless, oceanographers should not neglect bottle experiments in future work, as they provide perhaps the best way to probe the specific response of individual members of the plankton community⁷. On the other hand, mesoscale enrichments offer the chance to test the whole ecosystem response, but they are difficult to interpret because of unexpected or unquantifiable factors. Spin-offs of the 'iron hypothesis' — namely the depletion of important nutrients and carbon dioxide in surface waters — cannot be properly evaluated from the current results. The effects may be understood better after the follow-up iron fertilization experiments scheduled for next year.

The most resounding success of this complex experiment was to prove that it is possible to modify and then track a patch of open ocean surface water while measuring the ecosystem response. The experiment could hardly have been undertaken without the tireless effort and tenacity of John Martin. Regardless of the final judgement on the 'iron hypothesis', establishing the feasibility of such experiments may well turn out to be John's most enduring legacy. □

Mark L. Wells is at the Institute of Marine Sciences, University of California, Santa Cruz, California 95064, USA.

1. Martin, J. H. *Paleoceanography* **5**, 1–13 (1990).
2. Hart, T. J. *Discovery Rep.* **8**, 1–268 (1934).
3. Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
4. Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Nature* **345**, 156–158 (1990).
5. Ledwell, J. R., Watson, A. J. & Law, C. S. *Nature* **364**, 701–703 (1993).
6. Wells, M. L. *Biol. Oceanogr.* **6**, 463–476 (1988).
7. Price, N. M., Ahner, B. A. & Morel, F. M. M. *Limnol. Oceanogr.* (in the press).

Superficial gain

In the noble army of solids, metals form a regiment of their own. They owe their unique ductility, shine and conductivity to their special lattice of positive metal ions in a sea of negative electrons.

Daedalus points out that many non-metallic molecules, such as ammonium and the related alkylammoniums, also exist as stable positive ions. They might fit rather well into a metallic lattice. Indeed, ammonium forms an alloy with mercury. So Daedalus now plans to infiltrate various alien ions into the metallic lattice. By inserting ionic relatives of ammonium, ideally polymeric ones, he hopes to create a whole new class of organometallic 'hybrid alloys'.

His scheme is to electroplate the ions onto the metal from a suitable molten salt. Very high pressure may be needed to collapse the deposited ions into the dense metallic state so that they can alloy with the metal cathode and diffuse into it. The cathode will acquire a subtly graded composition: pure metal in the interior, but with a steadily increasing proportion of alloyed non-metallic ions towards the surface. Metal plated in this way should transform engineering.

Thanks to its loading of organic ions, a hybrid-alloy surface will be rust-proof, and will bind to paint and glue tenaciously. Its organic content will give it a degree of flexibility, making the surface layer slightly softer than the underlying metal. The resulting 'case softening', claims Daedalus, should make the metal extremely tough. For a loaded component usually fails from the surface inwards: the stress is greatest there, and small surface irregularities invite cracks to start. A case-softened surface would yield more easily, reducing the surface stresses. The stresses inside would be higher, of course; but the smooth inward gradation from surface alloy to pure bulk metal would give no sharp discontinuity from which a crack could begin. So, paradoxically, by softening its surface you would strengthen the whole component. It could be stressed right up to the elastic limit of the underlying metal without fear of brittle failure.

Daedalus also hopes to make homogeneous metal-organic hybrid alloys in bulk form, thus bridging the gap between metals and non-metals. Far lighter than most alloys, they should still retain that special metallic ability to be bashed and deformed enormously without breaking or losing strength. They might not be magnetic, but they should conduct electricity in intriguing ways. Indeed, they could even form a whole new range of cunning semiconductors.

David Jones

Antarctic benthic diversity

SIR — Poore and Wilson¹, and Rex *et al.*², give evidence for a distinct decline in species diversity in deep-sea communities from the tropics towards north polar regions; and for a less clear trend but high interregional variability in the Southern Hemisphere. Poore and Wilson¹ included some isopod data from the Scotia Basin and the Weddell Sea (Atlantic sector of the Southern Ocean), but the data of Rex *et al.*² are limited to regions north of 40° S.

We computed figures for species diversity by Hurlbert's method³ for Weddell Sea bivalves, gastropods and isopods, and find that these fall in the upper range of tropical values and are distinctly higher than in north polar regions (see table), provided that diversity figures depending on Agassiz trawl samples and epibenthic sled samples are comparable. Indeed, our data indicate that there is no steady latitudinal decrease in deep-sea benthic

range and of similar variability as in north polar seas^{7–9}. Quaternary glaciation in Antarctica was as extensive as in the north polar region¹⁰.

Hence, the tropical level of Antarctic species diversity is probably related to factors or processes unique to the Antarctic. The main differences between the Arctic and the Antarctic systems are the higher age and longer isolation of the latter. During the past 60 million years, the Antarctic experienced a slow and discontinuous transition from a warm water system in the early Tertiary (15 °C) to today's cold water system (–1.8 to 2 °C)¹¹. The first strong glaciation including shelf-ice formation had already occurred 36 million years ago in the early Oligocene¹². The formation of the circum-Antarctic current about 23.5 million years ago¹³ further isolated the Antarctic and its fauna from the world ocean. Most of the

logical record and present-day variation in species richness among taxa point to different levels of success in surviving recent periods of extreme glaciation¹⁴.

The high species diversity of the Antarctic slope and deep sea benthos shows the significance of regional-historical processes¹⁷ in the development of marine benthic communities. Thus, differences in regional history (geology, oceanography, climatology) have to be taken into account when interpreting global patterns of species diversity.

Thomas Brey, Michael Klages, Corinna Dahm, Matthias Gorny, Julian Gutt, Stefan Hahn, Michael Stiller, Wolf E. Arntz
 Alfred Wegener Institute for Polar- and Marine Research,
 Postfach 120161,
 D-27515 Bremerhaven, Germany
Johann-Wolfgang Wägele
Alfons Zimmermann
 Fakultät für Biologie,
 Universität Bielefeld,
 D-33615 Bielefeld, Germany

SPECIES DIVERSITY IN SOUTHEASTERN WEDDELL SEA

Taxon	No. of individuals	No. of species	$E(S_{100})$ in the Weddell Sea (this study)	$E(S_n)^*$ in the Weddell Sea (this study)	Range of $E(S_n)^*$ in tropical regions 20°N–20°S	Range of $E(S_n)^*$ in north polar regions ≥70°N
Bivalvia	190	26	20	17	3–16	<5
Gastropoda	162	34	27	18	10–23	<6
Isopoda	846	71	30	41	24–52	<10
Amphipoda	1,804	100	36	—	—	—
Decapoda	13,133	5	3	—	—	—
Holothuroidea	8,284	22	9	—	—	—
Ophiuroidea	4,031	23	13	—	—	—
Polynoidae	748	20	17	—	—	—

The table shows Agassiz trawl samples (500–2,000 m water depth, 70–78° S) compared with figures in ref. 2. For inter-taxon comparisons, we applied a normalized sample size of $n=100$.

* Normalized sample sizes used in ref. 2 are $n=75$ for bivalves, $n=50$ for gastropods and $n=200$ for isopods. $E(S_n)$, species diversity.

Quantitative social science

SIR — Double jeopardy (DJ) is a pervasive marketing phenomenon where a minor brand is often found to attract not only fewer buyers, but also ones that are less loyal. Ehrenberg¹ takes the generality of DJ in many consumer settings as a first demonstration of the existence of empirical "laws" with universal regularities in social sciences. He suggests that DJ conforms with an empirical model given by a Dirichlet distribution. We propose that such a model has a sound theoretical underpinning that stems from the basic laws of probability.

For simplicity, consider two competing brands with consumer preferences (degrees of loyalty) p and q such that $p + q = 1$ and p, q both ≥ 0 . Suppose there is a total of N consumers. The question is: how many of them will buy each brand? This is analogous to the classical coin-tossing experiment in which a loaded coin with biases p and q on both sides is tossed N times. The random outcomes N_p, N_q ($= N - N_p$) are given by the binomial distribution of the form $f(N_p | p) \equiv$

$$\binom{N}{N_p} p^{N_p} (1-p)^{N-N_p}.$$

The most likely outcomes, given by the peak of the distribution, are $N_p^* = N \cdot p$ and $N_q^* = N \cdot q$ so that $N_p^* > N_q^*$ if $p > q$. Thus a favourite brand necessarily yields a greater market share. Similar arguments are readily extended to situations with more than two brands or with multiple purchases from each buyer.

Conversely, suppose one observes that the market shares for two competing

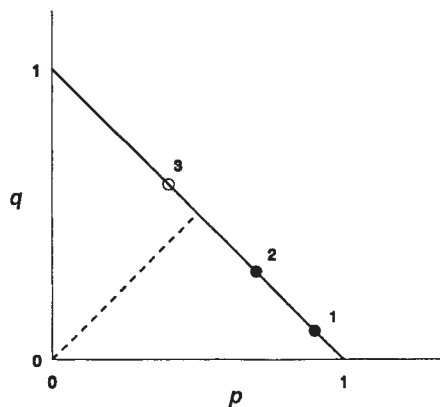
diversity towards south polar regions. But in the Weddell Sea species diversity varies by an order of magnitude among the taxa we analysed (see table). The polewards decline of species diversity in the Northern Hemisphere has been related to the inverse relation between magnitude and variability of sedimentation and species diversity², and to the effects of Quaternary glaciation on the benthos⁴. However, sedimentation of organic carbon in the southeastern Weddell Sea^{5,6} is in the same

present fauna evolved within the Southern Ocean¹⁴, as indicated by high levels of endemism¹⁵ exceeding those in the Arctic, where boreo-Arctic species dominate¹⁶.

The unique glaciological history of the Antarctic and its long term isolation led to the evolution of a specific fauna well adapted to a cold water system experiencing episodic glaciation and strong seasonal as well as inter-annual variability in food input (cold stenothermy, low metabolism and eurybathy). However, the palaeonto-

- Poore, G. C. B. & Wilson, G. D. F. *Nature* **361**, 597–598 (1993).
- Rex, M. A. *et al.* *Nature* **365**, 636–639 (1993).
- Hurlbert, S. H. *Ecology* **52**, 577–586 (1971).
- Dahl, E. *Sarsia* **64**, 57–59 (1979).
- Bathmann, U. *et al.* *Polar Biol.* **11**, 185–195 (1991).
- Wefer, G. & Fischer, G. *Mar. Chem.* **35**, 597 (1991).
- Bathmann, U. *et al.* *Prog. Oceanogr.* **24**, 117–129 (1990).
- Hebbeln, D. & Wefer, G. *Nature* **350**, 409–411 (1991).
- Bodungen, B. v. *et al.* in *Sediment trap in the Northern Countries* 2 (eds Wassmann, P. *et al.*) 116–136 (Yliopistopaino, Helsinki, 1991).
- Quilty, P. G. in *Antarctic Ecosystems—Ecological Change and Conservation* (eds Kerry, K. R. & Hempel, G.) 4–8 (Springer, Berlin, 1990).

- Kennett, J. P. J. *geophys. Res.* **82**, 3834–3860 (1977).
- Ehrmann, W. U. & Mackensen, A. *Palaeogeogr. Palaeoclim. Palaeoecol.* **93**, 85–112 (1992).
- Kennett, J. P. & Barker, B. F. *Proc. ODP Sci. Res.* **113**, 937–960 (1990).
- Clarke, A. in *Temperature and Evolution: Southern Ocean Cooling and the Antarctic Marine Fauna* (eds Kerry, K. R. & Hempel, G.) 9–22 (Springer, Berlin, 1990).
- White, M. G. in *Antarctic Ecology* Vol. 2 (ed. Laws, R. M.) 421–461 (Academic, London, 1984).
- George, R. Y. in *Polar Oceans* (ed. Dunbar, M. J.) 391–408 (Arctic Institute of North America, Montreal, 1977).
- Ricklefs, R. E. *Science* **235**, 167–171 (1987).



Probabilistic basis of double jeopardy. The preferences (p_i , q_i) of three buyers ($i = 1, 2, 3$) of two competing brands are defined by the linear constraint $p_i + q_i = 1$ (solid line) and satisfy $\sum_i p_i/3 = 2/3$ and $\sum_i q_i/3 = 1/3$. Buyers of major and minor brand (filled and open circles, respectively) are separated by the decision (dashed) line; buying decisions can be absolute (exclusive purchase) or relative (mixed) and are dependent on the values of p_i and q_i . It can be easily shown that $p_1 + p_2 = 1 + q_3$. Because $p_i \leq 1$ for all i , it follows that $q_3 \leq \min(p_1, p_2)$ and the equality holds only if $\max(p_1, p_2) = 1$. In practice, the above probabilistic description of DJ is valid only if market size is large.

brands are N_p and N_q . What can we say about consumer preferences based upon these observations? The answer is given by the conditional probability density in the form of a beta distribution² (a two-dimensional version of Dirichlet distribution) $\varphi(p|N_p) \equiv C \cdot p^{N_p-1} q^{N_q-1}$, where C is a function of N_p and N_q . The Bayesian mean estimates for consumer preferences are $p^* = N_p/N$ and $q^* = N_q/N$ so that $p^* > q^*$ if $N_p > N_q$. Thus greater market penetration implies greater consumer loyalty — quantity signifies quality.

As an example, suppose there are only three buyers whose average preferences for two brands are 2/3 and 1/3. From the binomial distribution the market composition is likely to be 2:1. The figure shows that except in extreme cases, the primary patron of the minor brand must necessarily be the least loyal. Thus DJ is dictated by consumer preferences and is completely predictable by the laws of probability.

There is no question that the social sciences can be as quantitative as any branches of science. The discovery of empirical laws is crucial, but mathematical modelling is indispensable if the benefits of the vast body of extant quantitative knowledge are to be realized.

Chi-Sang Poon

Harvard-MIT Division of Health Sciences and Technology,
MIT, Cambridge,
Massachusetts 02139, USA

Late Cretaceous mammals

SIR — Despite Madagascar's renowned subfossil fauna, and with the sole exceptions of a Miocene dugong¹ and an isolated, poorly dated Plio-Pleistocene faunule², this island-continent has until now yielded no confirmed pre-Holocene mammal fossils. We report here the discovery of a Late Cretaceous mammal from the Mahajanga (Majunga) Basin of northwestern Madagascar.

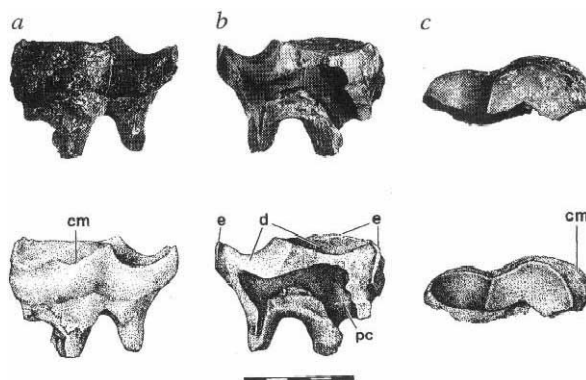
Late Cretaceous mammals from the Southern Hemisphere are known only from the Indian subcontinent³, which at that time lay adjacent to Madagascar and South America⁴; none has previously been recovered from Africa, Australia or Antarctica. The 30 or more million years of the Late Cretaceous is the interval during which the nontherian Multituberculata flourished and the diversification of marsupials and placentals began. However, despite the good record of Late Cretaceous mammals on northern continents⁵, the origins of many of the modern orders of placentals remain obscure; at least some probably originated on southern, Gondwanan-derived landmasses⁶.

In addition to numerous new records of lower vertebrates and invertebrates, we discovered a single specimen of a mammalian lower(?) cheek-tooth during a 1993 expedition to the Mahajanga Basin. The fossiliferous beds are believed to be of Campanian age⁷ and previously yielded only a moderately diverse assemblage of lower vertebrates, including dinosaurs⁸. We found the mammal specimen just east of Berivotra village, in a white unit at the top of the Maevarano Sandstone, below the succeeding Maastrichtian transgression. The specimen is incomplete and exhibits heavy premortem wear, but it is clearly that of a mammal as indicated by the presence of two roots, enamel that does not grade smoothly onto the root but terminates abruptly at the base of the crown, inflation of the crown beyond the dimensions of the roots, and the presence of a cingulum toward the mesial(?) end (see figure). The incomplete and worn nature of the tooth, as well as the paucity of knowledge of contemporaneous mammals from southern continents, precludes a precise identification. The most salient feature of the tooth is its size; it measures 7.55 mm in its longest dimension and is thus from one of the largest known Cretaceous mammals.

We think this specimen is a harbinger of things to come. Future discoveries of Late Cretaceous mammals in Madagascar will undoubtedly provide significant insights into the evolutionary history of mammals on southern continents and, specifically, of the highly endemic, imbalanced Madagascar fauna itself. The biographical origins of the biota of Madagascar, which started separating from the east coast of Africa in the middle Jurassic and reached its current position relative to the continent by the Early Cretaceous⁹, is one of the greatest unsolved mysteries of natural history.

David W. Krause*,

Department of Anatomical Sciences, State University of New York,
Stony Brook,
New York 11794, USA.



Fragmentary mammalian left (?) lower (?) cheektooth from the Upper Cretaceous Maevarano Sandstone, Mahajanga Basin, northwestern Madagascar in a, buccal view; b, lingual view of naturally fractured longitudinal section; and c, occlusal view. If a left lower tooth, mesial (anterior) is to the left in a and to the right in b and c. Abbreviations: cm, cingulum; d, dentine; e, enamel; pc, pulp cavity. Scale bar, 5 mm.

*Other authors of this letter: **Joseph H. Hartman**, EERC, University of North Dakota, Grand Forks, North Dakota 58202, USA; **Nell A. Wells**, Department of Geology, Kent State University, Kent, Ohio 44242, USA; **Gregory A. Buckley**, Department of Geology, Field Museum, Chicago, Illinois 60605, USA; **Charles A. Lockwood**, **Christine E. Wall**, **Roshna E. Wunderlich**, Department of Anthropological Sciences, State University of New York, Stony Brook, New York 11794, USA; **Joseph A. Rabarison**, **Louis L. Randriamaramanana**, Service de Paléontologie, Université d'Antananarivo, Antananarivo (101), Madagascar.

We thank B. Andriamihaja, B. Rakotosamimanana, B. Ramavivosa, E. Simons & P. Wright for logistical assistance in the field and the NSF SGER program for financial support.

1. Collignon, M. & Cottreau, J. *Ann. Paléont.* **16**, 135–171 (1927).
2. Sabatier, M. & Legendre, S. *Actes du 100^e Congrès national des Sociétés savantes, Montpellier, sciences*, fasc. VI, 21–28 (1985).
3. Prasad, G. V. R. & Sahni, A. *Nature* **332**, 638–640 (1988).
4. Bonaparte, J. F. *Natn. geogr. Res.* **6**, 63–93 (1990).
5. Lillegraven, J. A., Kielan-Jaworowska, Z., & Clemens, W. A. (eds) *Mesozoic Mammals: The First Two-Thirds of Mammalian History* (Univ. Calif. Press, Berkeley, 1979).
6. Krause, D. W. & Maas, M. C. *Geol. Soc. Am. Spec. Pap.* **243**, 71–105 (1990).
7. Besairie, H. *Ann. Géologiques de Madagascar* **30**, 1–463 (1972).
8. Sues, H.-D. J. *Paleont.* **54**, 954–962 (1980).
9. Rabinowitz, P. D., Coffin, M. F. & Falvey, D. *Science* **220**, 67–69 (1983).

1. Ehrenberg, A. *Nature* **365**, 385 (1993).
2. Freund, J. E. in *Mathematical Statistics* 2nd edn 281–282 (Prentice Hall, Englewood Cliffs, New Jersey, 1962).

Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary

Jim A. Williams, Stephen W. Paddock, Kathy Vorwerk & Sean B. Carroll*

Howard Hughes Medical Institute, Laboratory of Molecular Biology, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA

The appendages of arthropods and vertebrates possess a third, proximodistal patterning axis that is established after the primary anteroposterior and dorsoventral body axes by mechanisms that are largely unknown. The *vestigial* gene is required for formation of the entire *Drosophila* wing, and the dorsal/ventral boundary is shown to organize wing formation and *vestigial* gene expression. Interactions between dorsal and ventral cells in the growing imaginal disc induce *vestigial* gene expression through a discrete, extraordinarily conserved imaginal disc-specific enhancer. The link between dorsal/ventral compartmentalization and wing formation distinguishes the development of this sheet-like appendage from that of legs and antennae.

In *Drosophila*, thoracic segments contain ventral and dorsal sets of appendages. The dorsally located flight appendages (the wing and related haltere balancer organ) and ventral legs develop from imaginal discs formed during embryogenesis, grow during the larval stages of development, and differentiate during pupation. The developing imaginal discs are subdivided into anterior/posterior and dorsal/ventral compartments (in the wing) during embryogenesis and early larval development, respectively¹⁻⁴. The disc patterning fields also contain a proximal/distal (P/D) patterning axis and form distal (wing or leg) and proximal (body wall) structures (reviewed in ref. 5).

Mutational analysis has shown that the appendage-forming region within these discs requires the action of two classes of genes (reviewed in ref. 5). The first class is represented by the *decapentaplegic* (*dpp*)⁶, *wingless* (*wg*)^{7,8} and *aristaless* (*al*)⁹ genes, which are required to form distal regions within both leg and wing imaginal discs. The second class of genes is required specifically for P/D patterning in either dorsal or ventral discs but not both. The *Distal-less* (*Dll*) gene is a member of this class, because it is required to form a P/D axis in legs (and related ventral discs in cephalic segments) but it is not required in dorsal discs¹⁰. In contrast, loss of *apterous* (*ap*), *scalloped* (*sd*) or *vestigial* (*vg*) function results in total loss of distal derivatives of the wing but not the leg disc¹¹⁻¹³. As these dorsal or ventral disc patterning genes are required early in the development of the wing and leg fields, and loss of these genes specifically delete the entire appendage subregion without altering the remaining body wall portion of the field, these genes clearly function in the formation of the P/D axis of the field. Whether they act independently of or in concert with other factors necessary for patterning in these fields (such as *dpp*, *wg*) is unknown.

We have previously shown that wing formation requires the early action of the *wingless* and *apterous* genes and the formation of the dorsal/ventral boundary¹⁴. It is known that the dorsally restricted *apterous* homeodomain protein specifies dorsal cell fate¹⁵ and is required for D/V compartmentalization (refs 4, 15; and S. Blair, manuscript submitted). Here we demonstrate that one link between the organization of the imaginal disc into dorsal and ventral cell populations and wing formation is the activation of the *vestigial* gene. We show that interactions between cells of dorsal and ventral fate in the early wing disc induce

expression of the *vg* gene in a narrow band of cells spanning the D/V boundary by activating a newly identified, discrete regulatory element within the *vg* gene. The specific roles of *apterous* and *vestigial* in wing formation are key distinctions between the organization of the dorsal flight appendages and the cephalic and thoracic limbs.

Vg contains a D/V boundary response element

The requirement for *vg* and *sd* function to form distal wing disc derivatives tells us these genes are key members of the genetic program guiding wing formation. Patterned expression of both the *vg* and *sd* pro-wing genes in the developing wing disc is initially detected in a narrow stripe of cells symmetrically spanning the D/V boundary, shortly after this boundary is formed¹⁴. It would appear, then, that the *cis*-acting regulatory elements within these genes interpret the crude D/V positional information present within the early disc, and somehow refine this information into a narrow stripe of gene expression. Thus, analysis of the control elements of these genes should shed light on how this process occurs.

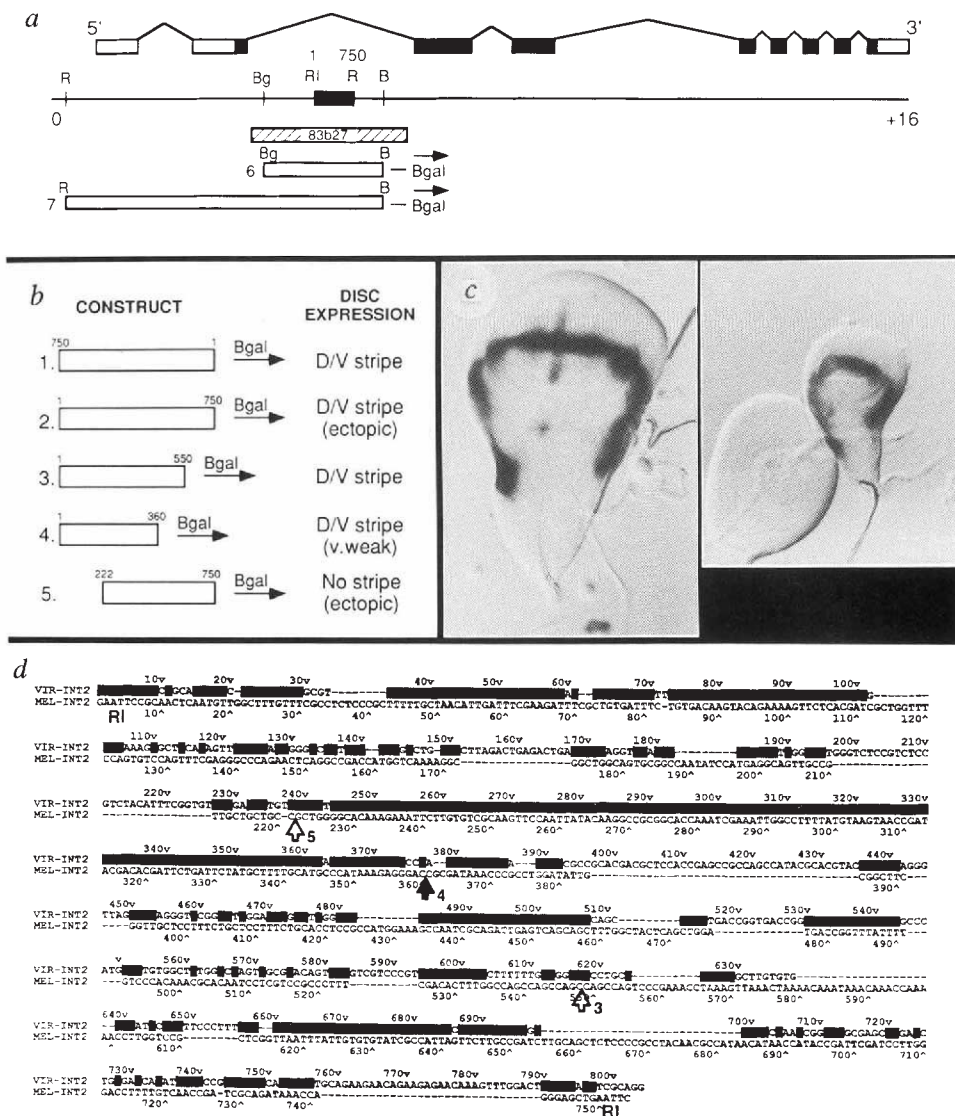
Both the *vg* and *sd* are large, and are spliced from exons spanning >16 kilobases (kb) of DNA. The location of wing disc control elements in the *sd* gene is unknown, but an allele of the *vg* locus (*vg*^{83b27}) provided a clue to the location of wing-specific control elements within this gene. No wings are formed in flies homozygous for this mutant, but this allele can complement other *vg* alleles in a way that is dependent on chromosome pairing (transvection)¹⁶. This type of genetic behaviour is predictive of regulatory mutations. The *vg*^{83b27} allele is associated with a deletion of most of intron 2 (Fig. 1a) and does not alter exons of the *vg* transcription unit¹⁷. When tested with an antibody raised against the *vg* protein, we find that early wing disc expression is lost in this mutant. This indicates that at least some wing disc control elements are located within intron 2 and demonstrates that these elements are essential to form distal wing structures. With this preliminary identification of the location of potential wing-disc-specific control sequences, we concentrated on this intron 2 region for further analysis.

Previous interspecific analysis of the DNA conservation patterns between the *D. melanogaster* and *D. virilis* *vg* genes had defined only one region with detectable DNA conservation within intron 2 (ref. 16; Fig. 1a). A β -galactosidase (β -gal) reporter gene containing this region (within a 750-base-pair (bp)

* To whom correspondence should be addressed.

FIG. 1 Identification of a D/V boundary response element within the *vg* gene. **a**, A partial restriction map of the 16-kb *vg* region with relevant *EcoRI* (R), (*Bgl*III (*Bg*) and *Bam*HI (B) sites and exons indicated. A 750-bp interspecifically conserved *EcoRI* fragment is shown in bold on the map, with orientation indicated from base pair 1 (promoter proximal) to 750 (promoter distal). The location of the *vg*^{83b27} deletion is depicted by a hatched box below the restriction map, and the position and orientation of β -gal reporter constructs 6 and 7 by open boxes and arrows. **b**, The size and orientation of five reporter constructs containing either the interspecifically conserved 750-bp *EcoRI* fragment (constructs 1 and 2) or three smaller DNA fragments from this region (constructs 3–5) are indicated on the left, and their respective imaginal disc expression patterns on the right. Expression is restricted to the wing and haltere discs in all cases. 'Ectopic' refers to β -gal expression in the notal and hinge regions of the wing disc which appears to originate from the dispensable 550–750-bp region and is probably an artefact arising from placing this region between the core enhancer and the target β -gal promoter in constructs 2 and 5. **c**, Late third instar imaginal discs from transformed flies containing one copy of reporter construct 1, stained with X-gal to reveal β -gal accumulation. Note that expression is restricted to the wing and haltere discs, and does not accumulate in the leg disc. **d**, DNA sequence of the 750-bp *EcoRI* fragment, and the corresponding region from *D. virilis*, is shown. The entire *D. melanogaster* sequence is shown, with conserved bases in *D. virilis* indicated by blocking. Sequence gaps in the alignments are denoted by dashes. Numbered arrows indicate the endpoints of reporter constructs 3, 4 and 5.

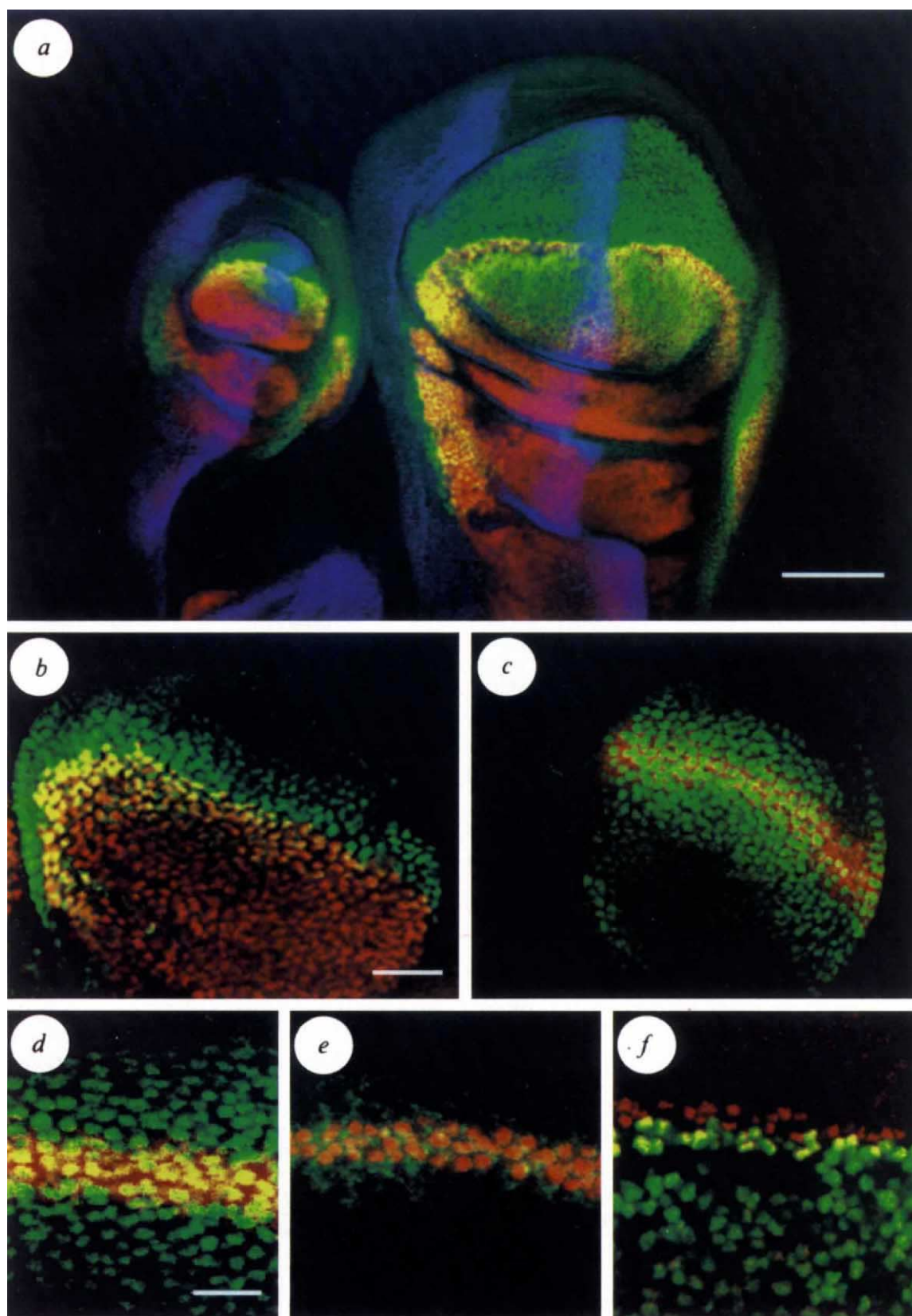
METHODS. The *D. melanogaster* *EcoRI* fragment, three selected deletion clones (Fig. 1b), and two larger *vg* region restriction fragments including intron 2 (Fig. 1a) were cloned in the indicated orientations (Fig. 1a and b) into the P-element transformation vector *hsp-lacZ*-CASPERS³² and used to transform *yellow white D. melanogaster* flies³³. Homozygous transformed lines from at least two independent trans-



forms were generated and analysed for each construct. X-gal staining was done as in ref. 14. The conserved intronic region in *D. virilis* was identified by hybridization of a random-primed ³²P-labelled probe of the 750-bp *D. melanogaster* intronic *EcoRI* fragment to Southern blots of restriction-enzyme-digested *D. virilis* *vg* region λ clones (λ V4 and V6; ref. 16) under low-stringency conditions. A single region of homology was detected, which was cloned, sequenced and aligned with the *D. melanogaster* intron-2 sequence (Fig. 1d).

EcoRI fragment) was constructed (reporter construct 1; Fig. 1b) and stable transformed lines were generated. Remarkably, β -gal expression driven by this element is restricted to a stripe of cells spanning the D/V boundary in the wing and related haltere discs (Fig. 1c), and no expression has been observed in any other larval or embryonic tissues, including ventral imaginal discs (for example, see leg disc in Fig. 1c). β -gal expression in early third instar wing discs is not reduced in *vg* or *sd* mutants (see below), indicating that expression is not the result of auto-regulation. Stable transformed lines of reporter constructs containing several smaller derivatives of this fragment were constructed (Fig. 1b), and a 360-bp minimal element capable of directing β -gal expression in a stripe spanning the D/V boundary was defined (reporter construct 4; Fig. 1b). This analysis has also revealed that this element can function independently of orientation (Fig. 1b, reporter constructs 1, 2 and 3) and

FIG. 2 The spatial expression domains of wing-patterning genes and the *vg* intron-2 enhancer element. **a**, The organization of the wing (right) and haltere (left) fields is similar as reflected by expression of: the *vg* protein (green) in the flight appendage primordia; the *ap* gene (red) in the dorsal compartment that bisects the *vg* pattern; and the *CiD* (blue) protein in all cells of anterior fate. **b–f**, The enhancer drives *lacZ* expression symmetrically on both sides of the D/V boundary. **b**, *ap* (red) bisects the native *vestigial* (green) protein pattern. Yellow and green cells represent the dorsal and ventral cells, respectively, of the future wing region. **c**, The enhancer drives *lacZ* (red: overlap with *vg* appears yellow) expression in a narrow band of cells that represents a subset of the native *vg* protein pattern (green). **d**, High-magnification view shows that this subset of cells is at the centre of the native *vg* domain which, as shown in **b**, is bisected by *ap* along the D/V boundary. **e**, The enhancer drives *lacZ* (green) in the same cells where the *cut* protein (red) is expressed along the D/V boundary in a late third instar disc; **f**, the *cut* protein (red) pattern is bisected by the domain of *ap* expression (green), demonstrating that the *cut*, and hence the enhancer pattern, straddles



► the D/V boundary. Scale bar: in a, 100 μm ; c, 20 μm ; d, 10 μm ; e, 20 μm .

METHODS. The primary antibodies used to detect $\beta\text{-gal}$ expression were either a $\beta\text{-gal}$ -specific mouse monoclonal antibody (Boehringer-Mannheim) or a rabbit anti- $\beta\text{-gal}$ polyclonal antibody. *CiD* expression was detected using a *CiD*-specific rat monoclonal antibody (kindly provided by C. Kelsey-Motzny and R. Holmgren). *vg* protein was detected using a *vg*-specific rabbit polyclonal antibody. These primary antibodies

were developed by two or three steps to either fluorescein or rhodamine conjugates using standard methodology¹⁶. Development and imaging of triply labelled wing discs, using rat (*CiD*), mouse ($\beta\text{-gal}$) and rabbit (*vg*) primaries to rhodamine, Cy-5, and fluorescein, respectively, are described in ref. 34. Protocols for dissection, antibody staining, and mounting of imaginal discs from second through to late third instar larvae have been described^{14,16}; the *ap/cut* image in Fig. 2f was provided by S. Blair¹⁸.

upstream (reporter constructs 1–4; Fig. 1b) or downstream (*vg* gene; Fig. 1a) of the target promoter. Thus, this control element acts as a true enhancer. The entire *EcoRI* fragment, and the corresponding region from *D. virilis*, were sequenced (Fig. 1d). The minimal element contains no obvious repeating motifs, and is remarkably conserved, as reflected by a region of 119 contiguous bases (and 135 out of 137) that are perfectly conserved between the two species. The extent of enhancer sequence conservation (which is the longest perfectly conserved region yet reported between these two species), the expression pattern that this DNA fragment directs, and the fact that loss of this region in the *vg*^{83b27} allele leads to total loss of distal derivatives from the wing disc, suggest that we have identified the key regulatory element within the *vg* gene that interprets the dynamic D/V patterning inputs in early wing and haltere discs and resolves it into a stripe of *vg* gene expression centred on the D/V boundary.

Activation of the *vg* control element

To determine the degree to which expression and regulation of the enhancer reflect the native *vg* gene, we examined the dynamics of enhancer-driven reporter gene expression in wild-type and mutant developing wing discs. For the remainder of this article we will focus on the larger wing disc, because all events in the two discs are similarly coordinated and the A/P, D/V and P/D organization of the smaller haltere disc is remarkably similar to that of the wing (Fig. 2a). β -gal expression driven by reporter construct 1 is not detectable in any tissues until approximately 8 h after the beginning of the second larval instar, at which time weak, diffuse expression is detected in broad domains within the wing. By the late second instar, when the native *vg* protein is expressed in a broad D/V stripe (bisected by the pattern of *ap* expression; Fig. 2b), the reporter gene is also expressed at high levels along the D/V boundary. The activation of the *vg* enhancer is the earliest marker of the D/V boundary of the wing future (*wg* and *cut* expression arise later along the wing margin; see below).

As the wing pouch and the native *vg* protein pattern expand in the early third instar, reporter gene expression is still driven in a narrow band of cells straddling the D/V boundary (Fig. 2c, d). Two larger expression constructs (6 and 7 in Fig. 1a) also exhibit expression along the D/V boundary of the wing pouch. Thus the intronic *cis*-regulatory element only drives *vg* at the patterning focus where dorsal and ventral cells meet. Other elements (outside intron 2; J.A.W., J. Esch and S.B.C., unpublished results) are responsible for the expansion of the *vg* pattern throughout the wing pouch. The enhancer continues to drive reporter gene expression throughout the third instar in cells on both sides of the D/V boundary, as demonstrated by its coincident pattern with the latter arising *cut* protein that marks the developing wing margin (Fig. 2e) and straddles the D/V boundary¹⁸ (Fig. 2f).

The expression of both the enhancer and *vg* protein in the wing region of late second instar, and both early and late third instar wing discs depends on formation of the D/V boundary. Gene expression in the wing region is completely abolished in *ap*^{null} and *wg*^{CX3}/*wg*^P (pupal lethal) genetic backgrounds that eliminate the D/V boundary (not shown). Temperature shifts from permissive to non-permissive temperatures of *wg*^{IL}/*wg*^{CX3} larvae (*wg*^{IL} is a temperature-sensitive *wg* allele) demonstrated that enhancer expression requires *wg* function before the third instar. In contrast, removal of *wg* function by temperature shifts after the early third instar did not affect enhancer expression (not shown). The timing of this *wg* requirement correlates well with the fact that *wg* is required to specify ventral identity during second instar development^{8,14}, and supports the argument that the *wg*-dependent specification of ventral identity is a prerequisite to wing formation. However, these observations do not explain why ventral specification or *ap* function are necessary to activate *vg* gene expression and hence, to promote wing forma-

tion. We next describe the results of two experiments using the *vg* control element that clarifies this relationship between D/V specification, *vg* gene activation, and wing formation.

Induction of *vg* expression

Given the dorsally restricted expression of *ap*^{14,15,18} and the requirement for *ap* function for the *vg* gene to be activated along the D/V boundary¹⁴, we sought to determine how *ap* expression and function is related to *vg* gene activation. We examined the most immediate effects of loss of *ap* function in small clones of imaginal disc cells. Using radiation to induce mitotic recombination, small clones of *ap* cells were generated in a genetic background containing reporter construct 1. The clones were marked in imaginal discs according to ref. 19, in which mutant clones are identified immunohistochemically by loss of a ubiquitously expressed *myc* epitope tag (Fig. 3 methods). Although *ap* function is necessary to form the entire wing region of the disc, small clones of *ap*⁻ tissue can survive throughout the wing disc. When *ap*⁻ (*myc*⁻) clones were induced in ventral cells, no alteration to enhancer expression was observed (Fig. 3a–c). However, when *ap*⁻ (*myc*⁻) clones were induced in dorsal cells, the striking observation was that the *vg* enhancer was activated in the *ap*⁻ cells bordering the clone boundary (Fig. 3d–m). This enhancer activation occurred even in very small clones (Fig. 3k–m) and was independent of position, arising in *ap*⁻ clones that arose in dorsal cells from disc regions fated to form either wing structures (Fig. 3h–m) or body wall structures (Fig. 3d–g). These results are consistent with the idea that *ap* is a key dorsal determinant and that loss of *ap* expression causes cells to switch from dorsal to a ventral-like fate (ref. 15, and S. Blair, submitted). Where cells of these two fates meet, whether in *ap*⁻ cells in *ap*⁺ fields, or at the wild-type D/V boundary, expression from the *vg* enhancer is induced.

Two signals for *vg* induction?

Whereas enhancer-driven β -gal expression along the D/V boundary is symmetrically distributed in both dorsal and ventral cells, enhancer expression induced by *ap*⁻ clones is predominantly in the *ap*⁻ cells (Fig. 3g, j, m). This asymmetry indicates that enhancer induction at the D/V boundary may involve at least two signals. One signal, putatively from dorsal to ventral cells, induces enhancer expression in ventral cells, and is accurately reflected in *ap*⁻ clones; this indicates that the ability to send this signal is controlled at the level of *ap* expression. The second potential signal, from ventral to dorsal cells, is not fully activated in *ap*⁻ clones, indicating that some additional property of ventral cells, other than absence of *ap* expression, is necessary fully to stimulate dorsal cells to activate the *vg* enhancer.

Wing formation and restricted *apterous* function

The loss of *wg* and *ap* function in early wing discs or of *ap* in clones eliminates or alters D/V compartmentalization and wing formation. The isolation of the *vg* D/V boundary response element allows us to examine the effect of disrupting this boundary in the developing wing disc. As the enhancer is activated on both sides of the D/V boundary, we would predict that ectopic expression of *apterous* driven by the *vg* enhancer would disrupt the position of the D/V boundary and would reduce the size of the wing field by progressively eliminating the ventral region. An *ap* cDNA was placed under the control of the *vg* enhancer and three independently transformed fly lines were produced which each yielded the predicted dose-dependent effect. With four copies of the transposon, most of the wing is reproducibly ablated (Fig. 4c), whereas two copies yields reproducible intermediate phenotype (Fig. 4b). Given that wing ablation occurs in a wild-type genetic background, this demonstrates that wing formation requires dorsally restricted *ap* expression, presumably to induce wing formation through interaction with ventral cells at the D/V boundary¹⁵.

Establishment of the patterning focus

Early third instar *vg* null wing discs are morphologically normal¹⁴, but the wing region of late third instar discs (identified morphologically as the wing pouch) does not develop¹⁴. Premature metamorphosis of wing discs from *vg* and wild type larvae has shown that early third instar *vg* discs are incapable of forming distal (wing) derivatives²⁰, demonstrating an early requirement for *vg* function in wing development. These results are consistent with *vg* functioning either to form the wing region or in its patterning. It has previously been impossible to distinguish

between these possibilities because expression of both the *vg* and *sd* genes (the only identified markers for the wing region in early discs) are co-dependent and lost in mutants for either gene¹⁴. However, the isolation of the *vg* enhancer, which also marks the wing region in early discs, now allows this question to be addressed. The enhancer was introduced into *vg* null and *sd* pupal lethal mutant backgrounds, and late second instar through to early third instar wing discs were stained for enhancer expression. Enhancer expression in *vg* or *sd* mutant discs was identical to that in wild-type discs. This demonstrates that the region that

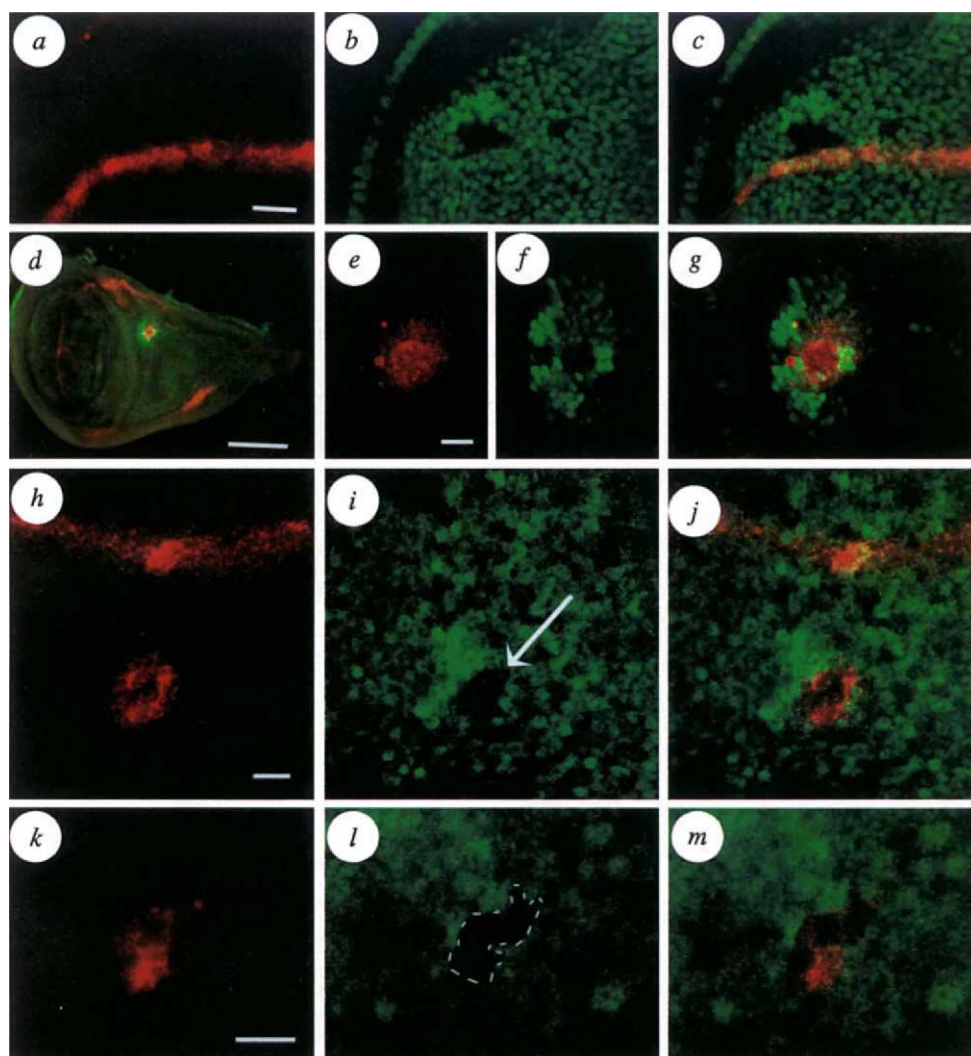


FIG. 3 Induction of a control element at the boundary of two cell populations. β -gal reporter gene expression is in red and the Myc epitope tag marking ap^+ cells is in green. The ap^- clones are detected by loss of the Myc tag; adjacent clusters of double-intensity Myc labelled cells are mitotic twin spots containing two copies of this tag (b, f, i and l). No enhancer activation is observed in ventral ap^- clones (a–c): ectopic expression of the enhancer is shown in three independent dorsal ap clones (d–m) including small clones (clone in k–m contains <10 cells, viewed at 4 $\times$ the magnification of the other panels). In all panels except d, the discs are oriented with ventral towards the top. A lower magnification view of a clone on the dorsal notum is shown in d, ventral is to the left. β -gal expression driven by reporter construct 1 (red: a, d, h and k) and expression of a ubiquitous Myc epitope tag (green; b, f, i and l) are merged (c, d, g, j and m) to show that activation of the enhancer is predominantly in the ap^- cells in dorsal clones (g, j and m). The bright red cells in k are in the ap^- clone (l, dotted outline). Note that little expression of the enhancer is observed in ap^+ cells that contact the boundary of dorsal clones of ap^- cells (g, j and m). Scale bars in a, e, h, k represent 10 μ m, and 100 μ m in d.

METHODS. Stocks of *ap* null and the heat-shock-inducible 42 π M *myc*

tag chromosomes were crossed to the *CyO/In(2LR)Gla, Bc; int1/int1* stock to create reporter construct 1 containing *In(2LR)Gla, Bc* balanced stocks of each chromosome. After genetically crossing the two stocks, staged larvae were collected, and a γ -ray source at 4 krad was used to induce mitotic recombination. After two days of growth at 25 degrees, *myc* expression was induced by heat shocking for 60–90 min at 37 $^{\circ}$ C. Wing discs from late third instar, non-*Bc* larvae, were dissected, fixed and stained with a Myc-specific monoclonal antibody (kindly provided by S. Blair) and β -gal-specific antibodies as described for Fig. 1. After mounting, mitotic recombination events were identified as clones of cells in which *myc* expression is lost. The presence of a bordering mitotic twin spot containing two copies of the Myc tag provides confirmation that the identified clone arose from a recombination event. Given that the *ap* gene is located at the base of the 2R chromosome (41B-C) and the 42 π M *myc* tag is at 45F¹⁹, two classes of *myc^-* clones are expected, depending on whether the identified recombination event is distal (*myc^-* clone) or proximal (*myc^- ap*) to the *ap* gene. We deduce that the enhancer-activating clones identified only in dorsal clones represent the ap^- subset (about one third of the total) of the *myc^-* clones.

would normally give rise to the wing is present in *vg* and *sd* mutant discs, even though these discs cannot form wing derivatives²⁰. The expression of the *vg* enhancer reflects a defined position within the disc that is subsequently set to a wing margin positional value by a process that is *vg*-dependent.

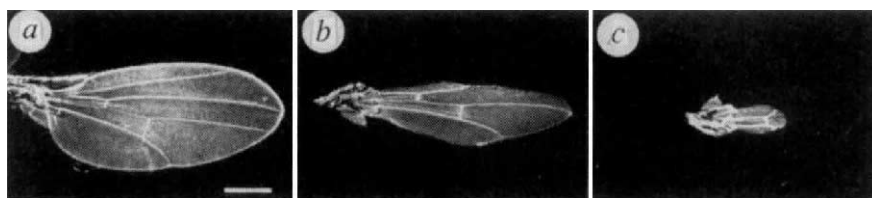
Induction event between two cell populations

It has been proposed that pattern within appendage fields may be organized from cell determination boundaries^{21,22}. Our results demonstrate that the dorsal/ventral compartment boundary of the wing imaginal disc is such an organizer. Once the growing disc is separated into dorsal and ventral cell populations, expression of the *vg* gene is induced at the boundary of the two cell populations, and development of the wing is set in motion (Fig. 5). Our results and recent clonal analysis¹⁵ indicate that the ability of dorsal and ventral cells to interact and promote this induction is controlled primarily at the level of *ap* expression. This gene is normally expressed only in wing disc cells of dorsal fate^{14,15,18}. Clonal analysis here has shown that induction of the *vg* enhancer occurs in clones of *ap*⁻ cells whenever they contact

dorsal cells. Thus, loss of *ap* expression is equivalent to a switch to ventral identity, as assayed by the ability to receive an activating signal from dorsal cells. These results strongly support the view that the difference between dorsal and ventral identity is controlled at the level of the *ap* gene product, which is a LIM-domain-containing homeodomain protein related to other proteins that function in the specification and/or maintenance of cell fate^{23,24}.

The lower level of enhancer activation in dorsal cells flanking *ap*⁻ clones suggests that the symmetrical induction event at the D/V boundary is the result of two signals, one from dorsal to ventral cells and another from ventral to dorsal cells (Fig. 5, step 3, inset). Additional factors, other than the absence of *ap*, appear to be necessary in ventral cells to fully signal dorsal cells to express the enhancer. It appears then that two induction events produce the symmetrical stripe of enhancer expression along the D/V boundary. Clearly the identification of the factors guiding cell-cell interactions and the regulation of the enhancer will be the key to understanding this induction process in detail. The size of the conserved region of the enhancer, the apparent

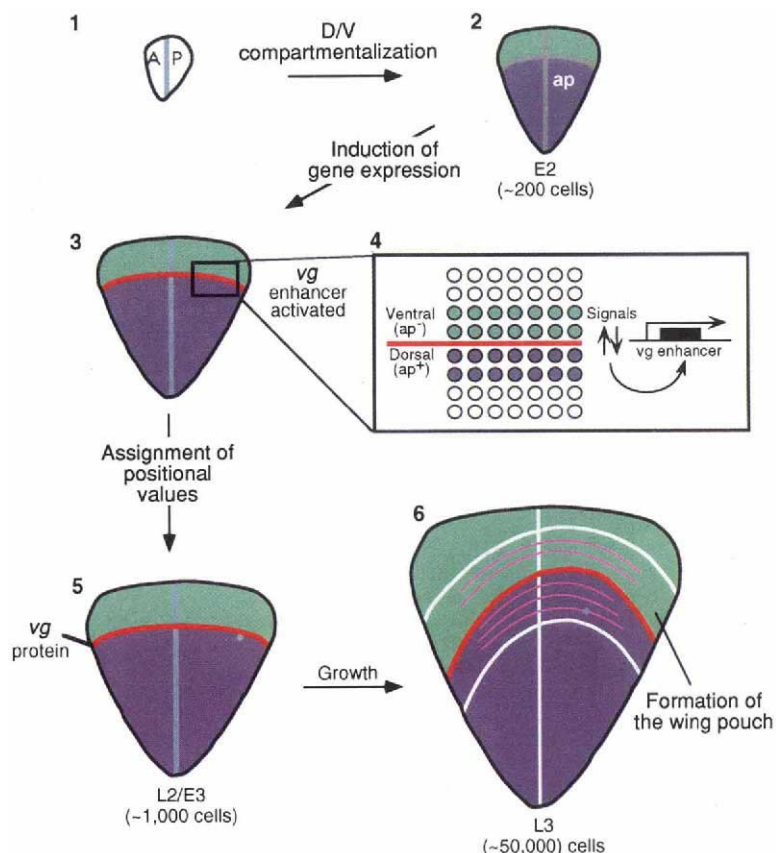
FIG. 4 Ectopic *ap* abolishes wing formation. The structure of a wild-type wing (a) is contrasted with wings from wild-type flies containing either 2 (b) or 4 (c) copies of a construct in which an *ap* cDNA is placed under the control of the *vg* intronic enhancer. All wings are from male flies. Scale bar, 0.5 mm.



METHODS. Expression of the *ap* coding region was driven by the *vg* enhancer in a modified CaSpeRhs vector (provided by C. Thurnmel). *Sall* linkers (New England Biolabs) were ligated to purified, blunt-ended, intronic *vg* *EcoRI* fragment DNA. After restriction with *Sall*, this fragment was cloned into *XhoI*-linearized CaSpeRhs vector, generating a recombinant vector in which the *vg* enhancer is immediately upstream of the heat-shock (*hs*) promoter. The coding region, contained in an 1.8-kb *XbaI*-*SmaI* restriction

fragment, of an *ap* cDNA (kindly provided by B. Cohen) was directionally cloned into the polylinker (restricted with *XbaI* and *StuI*) downstream of the *hs* promoter. A recombinant clone was used to transform yellow white *D. melanogaster* as described, and three independent transformed lines were isolated. Adult wings were dissected and mounted in Gary's magic mountant³⁰.

FIG. 5 D/V compartmentalization, induction of *vg* gene expression, and wing formation. An early larval wing disc that is divided into two cell populations of either anterior or posterior fate (step 1) begins to express the *wg* protein ventrally (step 2). *wg* establishes a ventral identity to these cells, and restricts *ap* expression to dorsal cells¹⁴. After establishment of D/V identity (step 3), cells at the D/V boundary interact through distinct signals to activate the *vg* enhancer in a stripe of cells symmetrically flanking this boundary (step 4). This results in *vg* protein accumulation that distalizes the positional value of this stripe of cells (step 5). The resulting positional discrepancy between this distal cell stripe and the adjacent proximal body wall cells stimulates wing growth by intercalation of wing blade positional values during subsequent disc development (step 6).



absence of repeating sequence motifs, and the genetic evidence for more than one signal at the boundary, suggest that the enhancer is receiving multiple regulatory inputs.

The cell interactions at the D/V boundary are analogous to previously characterized interactions involving *wg* and *en* at the A/P (parasegmental) boundary during embryogenesis (reviewed in ref. 25). In the latter case, anterior *wg*-expressing cells send a signal that maintains *en* expression in a two-cell-wide stripe of posterior cells; the posterior *en*-expressing cells at the A/P boundary send a signal back (hedgehog) that maintains *wg* expression in the flanking anterior stripe of cells. These two signals activate narrow stripes of gene expression flanking the boundary between two cell populations and maintain *gooseberry* gene expression in cells on both sides of the A/P boundary. In the wing disc, recent experiments suggest that signals across the A/P boundary play a key role in organizing the anteroposterior axis of the developing wing (K. Basler and G. Struhl, personal communication). The stability of the D/V boundary suggests that similar interactions between dorsal and ventral cells organize the D/V axis.

What does *vg* do?

Once the *vg* enhancer is activated along the D/V boundary, the *vg* protein is then required for the growth of the wing region. It is possible that *vg* regulates the positional value of this stripe of cells along the future edge or perimeter of the wing. Both *vg* and *sd* could function by assigning the positional values to the patterning focus by regulating one or more signalling pathways at the patterning focus. The *vg* protein is novel, but its nuclear localization does suggest a role in transcriptional regulation¹⁶. The *sd* protein is a TEA-domain DNA-binding protein, whose structure is highly conserved between invertebrates and vertebrates¹². The human *sd* cognate (*TEF1*) has been shown to require interactions with a tissue-specific protein to activate transcription of a target promoter²⁶. As expression of *vg* and *sd* in the wing is co-dependent¹⁴, it is possible that the *vg* and *sd*

proteins act together as transcriptional regulators of signalling pathways at the D/V boundary. The cells in this region may then act either as a source of a diffusible morphogen or to initiate a chain of sequential inductive signals emanating from the boundary. This could act to stimulate growth of the wing pouch and differentiation of pattern elements.

Patterning differences in legs and wings

On the basis of embryological^{27,28} and palaeontological evidence²⁹, it has been proposed that the dorsal disc patterning fields that give rise to the flight appendages of higher insects are derived from ancestral leg patterning fields (reviewed in ref. 5). Indeed, conservation of *wg*, *dpp* and *al* function during appendage formation in both dorsal and ventral discs⁹, as well as the fact that the *wg* gene is required to set ventral identity in both wings¹⁴ and legs^{8,30}, is consistent with this common origin theory⁵. However, wings are morphologically very distinct appendages, one major difference being the presence of a well defined edge (margin) and no obvious distal tip.

We suggest that the key evolutionary modification necessary to form and pattern a sheet-like (wing) rather than tube-like appendage (leg and antenna), was the formation of the D/V boundary. The leg disc has neither a well defined edge nor a definitive D/V boundary³¹, and P/D patterning within this disc may be directed primarily from a point focus at the centre^{9,10}. It is possible that the progenitor of the wing-patterning field acquired a D/V boundary, and that such a boundary was a necessary prerequisite to the evolution of a line focus, from which the sheet-like wing is formed. Indeed, we have discovered that *ap* gene expression is also dorsally restricted in butterfly wing discs (S.B.C. *et al.*, manuscript submitted), suggesting that the D/V compartmentalization and the wing formation programme described here is evolutionarily conserved, perhaps among all winged insects, and may indeed represent a fundamental developmental innovation in organizing a sheet-like appendage. □

Received 15 November 1993; accepted 11 February 1994.

- García-Bellido, A. *Ciba Found. Symp.* **29**, 161–182 (1975).
- Wieschaus, E. & Gehring, W. J. *Dev. Biol.* **50**, 249–263 (1976).
- Lawrence, P. A. & Morata, G. *Dev. Biol.* **50**, 321–337 (1976).
- Lawrence, P. A. & Morata, G. *Nature* **366**, 305–306 (1993).
- Williams, J. A. & Carroll, S. B. *Bioessays* **15**, 567–758 (1993).
- Spencer, F. A., Hoffman, M. & Gelbart, W. M. *Cell* **28**, 451–461 (1982).
- Baker, N. E. *Development* **102**, 489–497 (1988).
- Couso, J. P., Bate, M. & Martínez Arias, A. *Science* **259**, 484–489 (1993).
- Campbell, G., Weaver, T. & Tomlinson, A. *Cell* **74**, 1113–1123 (1993).
- Cohen, S. M. in *Development of Drosophila* (eds Martínez Arias, A. & Bate, M.) (Cold Spring Harbor Press, New York, in the press).
- Cohen, B., McGriffin, M. E., Pfeifle, C., Segal, D. & Cohen, S. M. *Genes Dev.* **6**, 715–729 (1992).
- Campbell, S. *et al.* *Genes Dev.* **6**, 367–379 (1992).
- Williams, J. A. & Bell, J. B. *EMBO J.* **7**, 1355–1363 (1988).
- Williams, J. A., Paddock, S. W. & Carroll, S. B. *Development* **117**, 571–584 (1993).
- Dias-Benjumea, F. & Cohen, S. M. *Cell* **75**, 741–752 (1993).
- Williams, J. A., Beil, J. B. & Carroll, S. B. *Genes Dev.* **5**, 2481–2495 (1991).
- Williams, J. A., Atkin, A. L. & Bell, J. B. *Molec. gen. Genet.* **221**, 8–16 (1990).
- Blair, S. *Development* **119**, 339–351 (1993).
- Xu, T. & Rubin, G. M. *Development* **117**, 1223–1237 (1993).
- Bownes, M. & Roberts, S. *Differentiation* **18**, 89–96 (1981).

- Meinhardt, H. *Dev. Biol.* **96**, 375–385 (1983).
- Meinhardt, H. *Sem. dev. Biol.* **2**, 129–138 (1991).
- Frey, G., Kim, S. K. & Horvitz, H. R. *Nature* **344**, 876–879 (1990).
- Way, J. C. & Chalfie, M. *Cell* **54**, 5–16 (1988).
- Ingham, P. W. & Martínez Arias, A. *Cell* **68**, 221–235 (1992).
- Xiao, J. H., Davidson, I., Matthes, H., Garnier, J. M. & Chambon, P. *Cell* **65**, 551–568 (1991).
- Anderson, D. T. *J. Embryol. exp. Morph.* **11**, 339–351 (1963).
- Cohen, B., Simcox, A. A. & Cohen, S. M. *Development* **117**, 597–608 (1993).
- Kukalova-Peck, J. J. *Morphol.* **156**, 53–126 (1978).
- Struhl, G. & Basler, K. *Cell* **72**, 527–540 (1993).
- Steiner, E. *Wilhelm Roux Arch. dev. Biol.* **180**, 9–30 (1976).
- Nelson, H. B. & Laughon, A. *Wilhelm Roux Arch. dev. Biol.* **202**, 341–354 (1993).
- Spradling, A. C. in *Drosophila, A Practical Approach* (ed. Roberts, D. B.) 175–197 (IRC, Oxford, 1986).
- Paddock, S. W., Langeland, J. A., DeVries, P. & Carroll, S. B. *BioTechniques* **14**, 42–48 (1992).

ACKNOWLEDGEMENTS. We thank Y. Xu and G. Rubin for stocks bearing the *myc* epitope tag; C. Kelsey-Molcay, R. Holmgren, S. Blair, A. Martínez Arias and R. Nusse for antibodies; S. Cohen, S. Blair and A. Chovnick for selected fly stocks; S. Blair for the image in Fig. 2f; F. Dias-Benjumea and S. Cohen for communication of results before publication; our colleagues for their comments on this work; and J. Wilson for preparing the manuscript and L. Olds for illustrations. This work was supported by an MRC postdoctoral fellowship to J.A.W. and by the Howard Hughes Medical Institute.

Structure of the regulatory domain of scallop myosin at 2.8 Å resolution

X. Xie^{*}, D. H. Harrison^{*}, I. Schlichting[†], R. M. Sweet[‡], V. N. Kalabokis[§], A. G. Szent-Györgyi[§] & C. Cohen^{*||}

^{*} Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254-9110, USA

[†] Max Planck Institut für Medizinische Forschung, D-6900 Heidelberg, Germany

[‡] Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

[§] Department of Biology, Brandeis University, Waltham, Massachusetts 02254-9110, USA

The regulatory domain of scallop myosin is a three-chain protein complex that switches on this motor in response to Ca^{2+} binding. This domain has been crystallized and the structure solved to 2.8 Å resolution. Side-chain interactions link the two light chains in tandem to adjacent segments of the heavy chain bearing the IQ-sequence motif. The Ca^{2+} -binding site is a novel EF-hand motif on the essential light chain and is stabilized by linkages involving the heavy chain and both light chains, accounting for the requirement of all three chains for Ca^{2+} binding and regulation in the intact myosin molecule.

In contrast to the myosins of vertebrate skeletal muscle, molluscan myosins are Ca^{2+} -regulated molecules¹⁻³. As in other muscle myosins, the N-terminal ~850 residues of the heavy chain are folded into a globular 'head' portion (subfragment-1, S1) (Fig. 1) which contains the ATPase and actin-binding sites of the molecule and also binds one each of two different light chains. These so-called 'regulatory' and 'essential' light-chain (RLC and ELC respectively) subunits are critical for regulation^{4,5}. Scallop myosin is especially suitable for studying the light chains because their removal or exchange can be easily achieved. Scallop myosin lacking the RLC is unregulated: it no longer binds Ca^{2+} ions and has a high ATPase activity independent of the Ca^{2+} concentration. The RLC inhibits activity at low levels of Ca^{2+} . The ELC has been inferred to be the site of specific Ca^{2+} binding, but its specificity and affinity depend on interactions of the ELC both with the RLC and with the heavy chain⁶. A stable proteolytic fragment of the scallop myosin head has been isolated which contains part of the heavy chain ($M_r \sim 10,000$ (10K)), together with the ELC and RLC (Fig. 1); this so-called regulatory domain has the Ca^{2+} -binding characteristics of intact myosin^{6,7}.

We report the crystallization and structure determination to 2.8 Å resolution of the scallop myosin regulatory domain with bound Ca^{2+} . The results reveal the detailed folding and side-chain interactions of the three chains in the complex, and how they generate an unusual Ca^{2+} -binding site. The structure also explains results of hybridization and mutagenesis studies on the light chains of this domain⁸. Our myosin fragment structure determination reveals the side-chain interactions critical for function, thereby complementing and extending the 2.8 Å crystal structure of the S1 fragment of chicken skeletal myosin⁹. In that study the presence of two ELC isoforms also prevented clear tracing of the main chain in some regions of this light chain, and virtually all the lysines were methylated. Some features of the structures are compared below, but detailed comparison awaits the availability of coordinates. Moreover, chicken myosin—like all vertebrate striated muscle myosins—is not a regulated molecule and does not have a Ca^{2+} -specific binding site. Our structural study of the regulatory domain of scallop myosin is thus a first step in understanding how information is transmitted from this distal domain of the myosin head to switch on the activity of the motor domain.

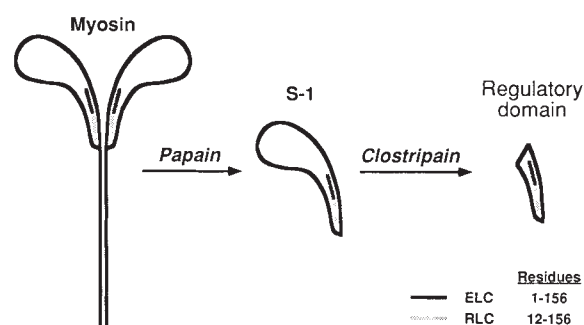


FIG. 1 Production of the scallop regulatory domain (RD) by proteolysis. The RD was prepared by digestion of papain-digested S1 with clostripain as described⁶. The procedure for the preparation of S1 (ref. 34) was modified to avoid overdigestion of S1. Papain (Boehringer-Mannheim) was purified by affinity chromatography³⁵. Eleven residues from the N-terminus of the RLC have been removed by papain.

Structure determination

Crystals of the complex were obtained from polyethylene glycol in the presence of 0.1 mM Ca^{2+} (Table 1 legend), and the structure was solved at 2.8 Å resolution. The crystals have the symmetry of the monoclinic space group $P2_1$ ($a = 52.5$, $b = 87.0$ and $c = 55.5$ Å; $\beta = 114.5^\circ$). There is one molecule per asymmetric unit; the solvent content of the crystal is ~54%. The phases were determined by multiple isomorphous replacement (MIR) combined with anomalous scattering (MIRAS) (Table 1). The final atomic model, refined at 2.8 Å resolution, has an R factor of 0.201 and good stereochemistry (Table 1). The quality of the final electron density map is shown in Fig. 2a.

Overall topology

The complex has a striking appearance: as in the chicken S1 structure⁹, the heavy chain forms a long α -helix with a sharp bend near the C terminus, and is stabilized throughout its length by the two light chains arranged in tandem (Fig. 2b). The overall dimensions of the complex are $\sim 100 \times 30 \times 35$ Å³; 67% of the residues are in the α -helical conformation. Both light chains have a dumbbell-like fold similar to that of calmodulin (CaM). The ELC winds around the N-terminal portion of the heavy-

|| To whom correspondence should be addressed.

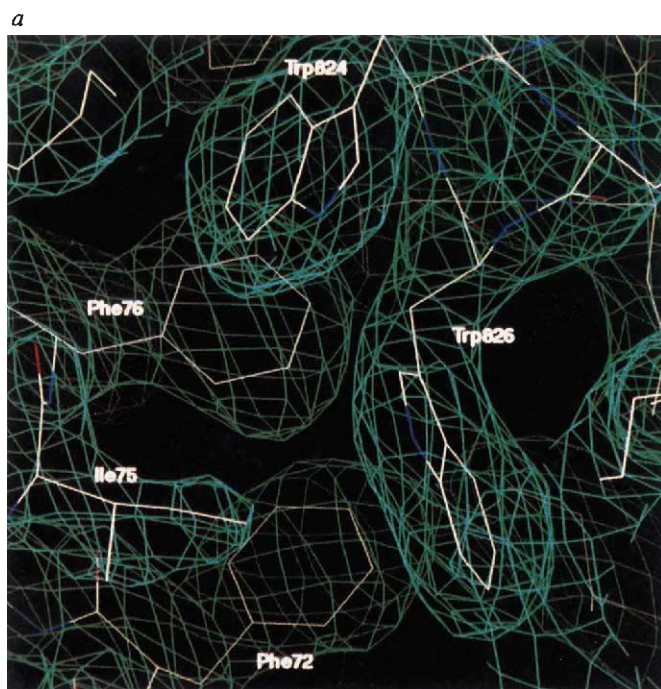
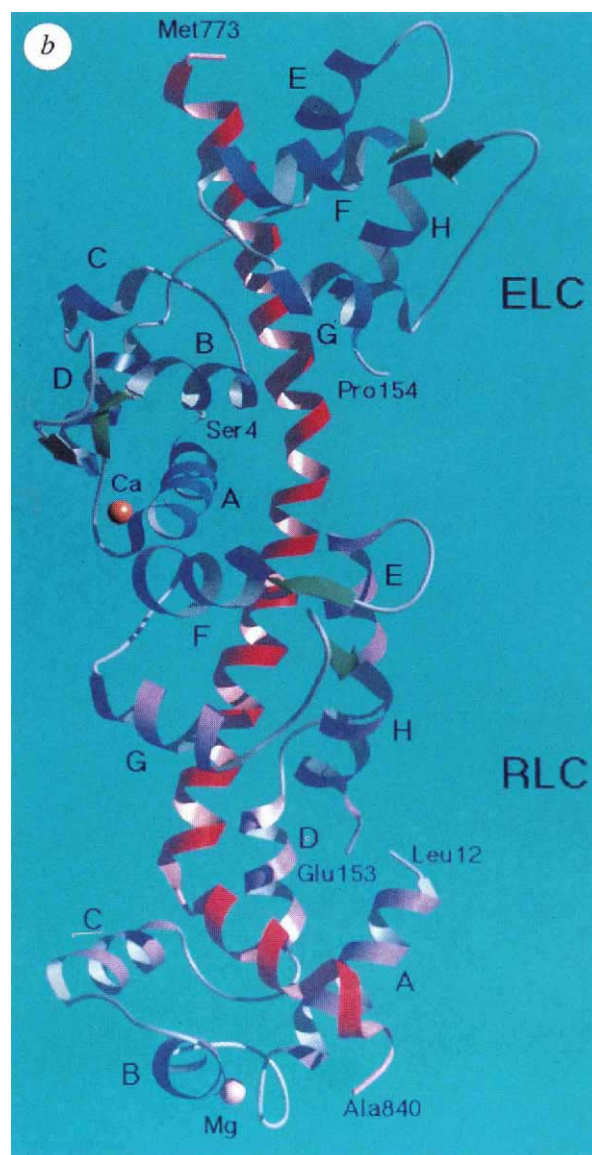


FIG. 2 *a*, The 2.8 Å electron density map of a portion of the RD complex, calculated with coefficients $(2|F_{\text{obs}}| - |F_{\text{calc}}|)$ and α_c phases from the refined 2.8 Å resolution structure, contoured at 1σ . The heavy-chain residues shown are Trp 824, Trp 826; those from the RLC are Phe 72, Ile 75 and Phe 76. In addition to the 11 residues of the N terminus of the RLC removed by cleavage with papain, there is no electron density for the last three C-terminal residues of this light chain nor for three residues at the N terminus and for two at the C terminus of the ELC. There are no breaks in the chain trace, and we can identify all of the side chains. At this resolution, the orientations of some side chains, as well as a number of the backbone carbonyl oxygen atoms, are uncertain. *b*, Ribbon diagram of the overall fold of the RD complex. The present model consists of 68 amino acids of the heavy chain (Met 773–Ala 840), 142 amino acids of the RLC (Leu 12–Glu 153), 151 amino acids of the ELC (Ser 4–Pro 154) and two metal ions. Heavy-chain residues are coloured red; ELC residues are in blue and RLC residues are in purple. The specific Ca^{2+} ion is shown in orange and the non specific divalent cation in pink. The nomenclature for the arrangement of the EF-hand domains is as follows: domain I begins with helix A, followed by a loop, then helix B followed by a linker; domain II begins with helix C, loop, then helix D, and so on. The loops are coloured grey, and the short regions of antiparallel β -sheet at the ends of loops I, II and III, IV are coloured green. (Figure prepared using graphics program Rayshade; E. C. Kolb, personal communication.) Each of the two lobes of the RLC, consisting of a pair of helix-loop-helix (EF hand) domains joined by a loop, interacts with adjacent segments of the C-terminal region of the heavy chain in a similar manner. The N-terminal lobe of the RLC grips the C-terminal 'hook' section of the heavy chain, and the C-terminal lobe grips the straight helix above in a similar fashion. The linker helix between lobes of the RLC is unwound slightly near Gly 82, thus allowing the equivalent positioning of the two lobes with respect to the heavy chain. Within each lobe the pair of EF-hand domains is related by an approximate local two-fold rotation axis. Although the ELC has a similar



pair of lobes connected by a linker region, their precise fold and interactions with the heavy chain differ somewhat from those of the RLC. The hydrophobic core of the C-terminal lobe of the ELC (formed by EF hands III and IV) grips the N-terminal portion of the heavy chain in the same way as do the two lobes of the RLC. But the N-terminal lobe of the ELC swings around so that only helices A and B of EF-hand I make contact with the edge of the heavy chain on the side opposite to that gripped by the C-terminal lobe. The helix in the linker region is thus considerably more unwound in this light chain than in the case of the RLC. Moreover, the helices flanking the two loops in the N-terminal domain are more antiparallel than perpendicular to one another and form a kind of 4-helix bundle, rather than an 'apolar channel' as in the other domains.

chain fragment and the RLC winds around the C-terminal region; both have a polarity opposite to that of the heavy chain. The two light chains make contact with one another over a very limited region. This contact area involves the first loop of the ELC, which is the specific Ca^{2+} -binding or 'triggering' site responsible for myosin activation. The nonspecific divalent-cation-binding site, which is occupied by a Mg^{2+} or Ca^{2+} ion, is located in the first loop of the RLC.

Notable features in the fold of the heavy chain are the $\sim 40^\circ$ bend in the middle portion of the long α -helix and the sharp

turn of $\sim 60^\circ$ forming a 'hook' towards the C terminus. The bend occurs over a limited number of residues (Arg 795, Lys 796, Ala 797, Tyr 798, Lys 799 and Lys 800) near the contact region between the ELC and RLC on the concave side of the heavy chain. The second, sharper turn in the heavy chain occurs at residues Trp 824, Gln 825, Trp 826; the chain continues as another short stretch of helix. This sharp turn positions all the hydrophobic residues in this segment of the heavy chain in the hydrophobic core formed by the RLC.

Except for the N-terminal lobe of the ELC, the contacts made

by the light chains with the heavy chain are somewhat similar to those found in CaM-peptide complexes^{10,11}, but in contrast to the peptides, the heavy chain is more exposed to solvent in certain regions along its length (see Fig. 2*b* legend, and below). As in the CaM-peptide case, the major contacts between the light chains and the heavy chain are hydrophobic (see later).

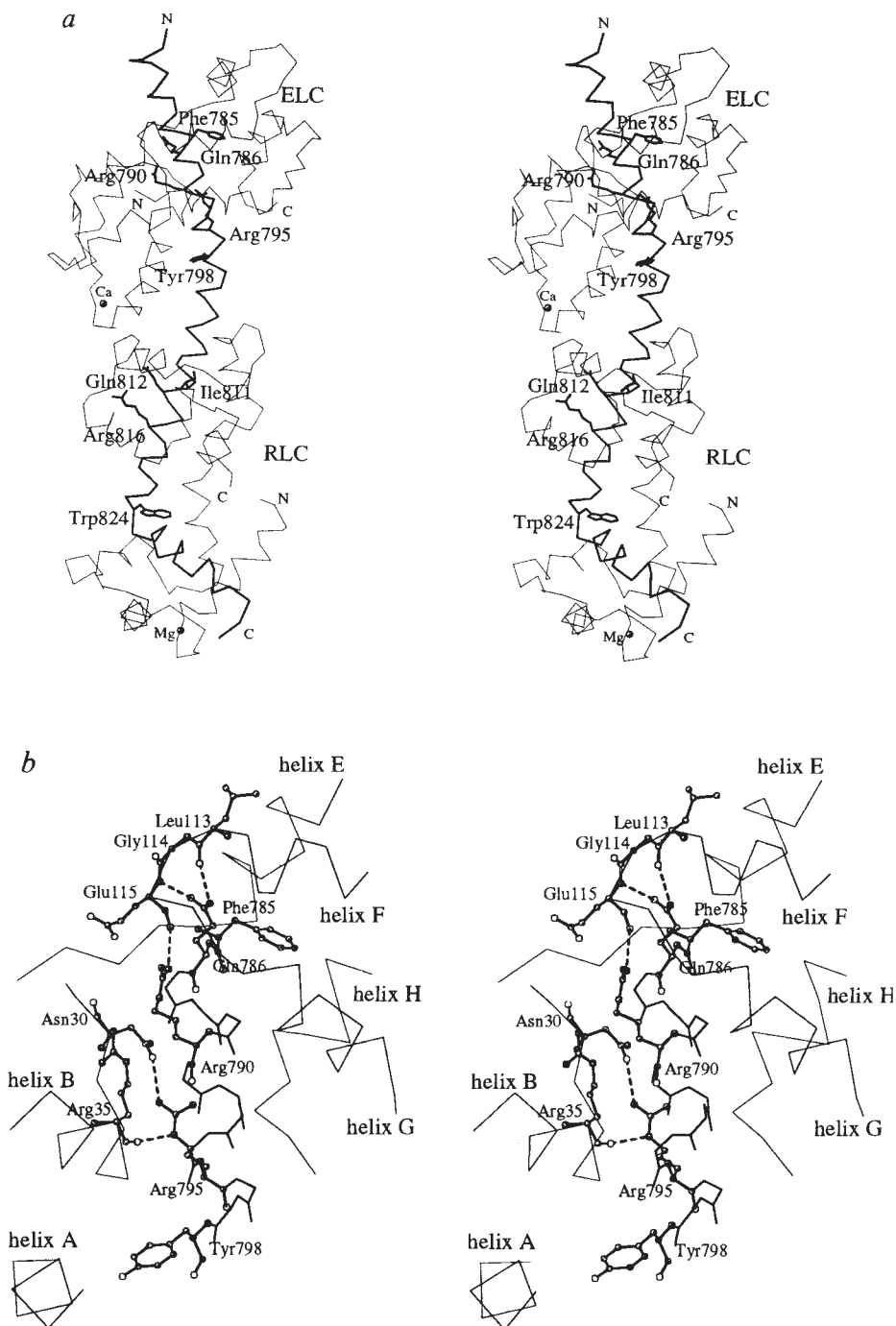
The IQ motif

There are 413 contacts (≤ 4 Å) (209 with the ELC and 204 with the RLC) between the heavy chain and the light chains. Of these contacts, 9% are hydrogen bonds and 88% are van der Waals contacts. The heavy chain is highly hydrophobic: about 50% of the residues are apolar and most of these are in contact with the light chains. Of the 19 charged residues in the heavy chain, only three are acidic and none of these is involved in light-chain

binding. In contrast, 7 out of 16 basic residues make hydrogen bonds and/or salt bridges with the light chains.

The so-called IQ motif (~23–25 residues), with a consensus sequence of the form IQxxxRGxxxR (in one-letter amino-acid code, where x is any amino acid), has been found in the heavy chains of many myosins, and signals the potential for calmodulin/light-chain binding^{12,13}. In the case of the scallop heavy chain, the first IQ motif (residues 785–795: Phe-Gln-Ala-His-Ile-Arg-Gly-Tyr-Leu-Ile-Arg) related to ELC binding follows the consensus pattern very well (Fig. 3*a*). The motif in the RLC binding zone (residues 811–821: Ile-Gln-Arg-Asn-Ile-Arg-Lys-Trp-Leu-Val-Leu) is less conserved. In both cases, however, the first three consensus residues make almost identical contacts with the light chains (Fig. 3*a, b*), strengthening the grip of the C-terminal lobe of the light chains to the heavy chain. The distribu-

FIG. 3 *a*, Stereoscopic view of the α -carbon backbone of the RD complex, traced using the program Molscript³⁶. Some of the key side chains of residues in the two IQ motifs of the heavy chain have been labelled so that the similar interactions made by each motif with the corresponding light chain can be seen. *b*, Stereoscopic view of interactions of the first IQ motif with the ELC. Phe 785 (and Ile 811 in the case of the RLC) fits into the hydrophobic core formed by the C-terminal lobe of the ELC. Moreover, the side-chain atoms of Gln 786 make hydrogen bonds with the main-chain carbonyl oxygen and amide nitrogen atoms at the end of helix F, and with the linker between helices F and G of the ELC. The side chain atoms of Gln 812 make similar contacts with the corresponding regions of the RLC. The residues of the two light chains that are hydrogen-bonded to the Gln residues of the heavy chain are very similar: Leu 113, Gly 114, Glu 115 in the case of ELC; and Met 116, Gly 117, Asp 118 for the RLC. The third key residue, Arg 790 (or Arg 816 for the RLC) of the IQ motif forms hydrogen bonds with residues in the same linker between helices F and G. The fifth key residue (Arg 795) forms hydrogen bonds with the linker between the B and C helix, as well as with the B helix of the ELC. This Arg residue is not present in the RLC binding region. In both light chains, the residues at the end of helix F, as well as the hydrophobic residues in helices F and G, are very similar to the corresponding residues in CaM. Note that the arrangement of the two EF-hand domains in each lobe is different from that in the CaM-peptide complexes. Proceeding from the N to the C terminus, the heavy chain is gripped by the light chain EF-hand domains as follows: III(ELC), IV(ELC), I(ELC), III(RLC), IV(RLC), II(RLC), I(RLC). In contrast, in the case of the CaM-peptide complex, the MLCK peptide is gripped from the N to C terminus by the EF-hand domains IV, III, I and II. This latter pattern is also found in the crystal packing of troponin C, in which the holo-C domain is bound to the A helix of the apo-N domain in a similar fashion³⁷.



tion of the key residues in the IQ motif requires an α -helical conformation for the heavy chain so that the appropriate bonds with the light chains can be formed (Fig. 3b). Thus there appears to be a clear structural basis for the binding of the CaM-light chain superfamily of proteins to the IQ motif of myosin heavy chains.

Sequence comparison has also indicated that a number of proteins or peptides that bind CaM with high affinity do not have the IQ motif, but share the common feature of two aromatic or long-chain hydrophobic amino acids separated by a stretch of

12 residues¹⁰. This feature is also present in the scallop heavy chain. In the case of the ELC binding site, the 14 residues begin at Phe 785 of the IQ motif and end at Tyr 798; in the case of the RLC, they begin at Ile 811 and end at Trp 824 (Fig. 3a). This sequence pattern positions these bulky hydrophobic residues about 3.6 α -helical turns apart, and allows the N- and C-terminal lobes of CaM or the light chains to make hydrophobic contacts with their target peptide or heavy chain from opposite sides in a similar fashion. Thus we would extend the consensus sequence for the IQ motif of myosin heavy-chain sequences by

TABLE 1 Statistics of crystallographic and refinement data

Parameter	Native	PIP	EMP	Pb	U
Resolution limit (Å)	2.8	3.3	3.0	3.5	3.7
Wavelength (Å)	1.150	1.067	1.004	0.945	1.150
Reflections (unique/observed)	10,352/30,711	4,984/12,289	6,223/14,739	5,302/9,854	4,333/5,706
($I > \sigma(I)$)					
R_{merge}^*	0.113	0.097	0.112	0.104	0.116
Completeness ($I > \sigma(I)$)	0.91	0.63	0.77	0.40	0.42
MIR statistics					
Soaking concentration (mM)		0.1	0.1	0.1	0.1
Soaking time (h)		6	5	5	4
Number of sites		2	3	3	2
Phasing power†		1.55	1.67	0.87	1.67
$R_{\text{iso}}^\ddagger$		0.19	0.29	0.15	0.41
Refinement statistics					
Resolution (Å)	10.0–2.80				
R factor§	0.201				
Reflections ($ F > 2\sigma(F)$)	8,960				
Completeness	0.81				
Number of RD atoms	2,938				
Number of Ca^{2+} atoms	2				
R.m.s. deviation from ideal bond distance (Å)	0.016				
R.m.s. deviation from ideal angle (°)	3.76				
Mean B factor (Å ²)	HC	RLC	ELC		
All atoms	29	32	32		
Backbone atoms	27	30	29		

The scallop regulatory domain consists of ~10K of the heavy chain (beginning at Leu 755 or Val 760 and continuing past Pro 834), together with the intact essential light chain (156 residues), and 145 out of 156 residues of the regulatory light chain⁶. The complex was crystallized in the presence of Ca^{2+} by the hanging drop method at 4 °C. The drops initially contained 10 mg ml⁻¹ protein in 2 mM MgCl_2 , 0.1 mM CaCl_2 , 75 mM HEPES (pH 7.0) and 7–8% (w/v) polyethylene glycol (PEG) 4000. Typically, crystals grew within three weeks to sizes up to $0.1 \times 0.3 \times 0.7$ mm³. The wells contained 1 ml of the same buffer and 14–16% (w/v) PEG 4000. All data sets were collected at 4 °C using an Enraf-Nonius FAST detector and synchrotron X-radiation at Brookhaven National Laboratory. Possible derivatives were identified by X-ray fluorescence measurements and derivative data were then collected at the appropriate X-ray wavelength to maximize the anomalous scattering signal. Phasing at 3.3 Å resolution commenced with the location of the two major platinum sites in the PIP derivative in difference Patterson maps. Additional heavy-atom sites were determined from difference Patterson and/or difference Fourier maps. Initially, the position and occupancy of the heavy atoms were refined against both isomorphous and anomalous differences for each derivative independently. Phase angles were then calculated from all derivatives and improved by solvent flattening³⁰. These unbiased phased angles were then used, together with isomorphous and anomalous differences, to refine the position, occupancy and B factor of all the heavy atoms together, giving an overall figure of merit of 0.57 for 6,835 reflections. The crystallographic package PROTSYS (G. A. Petsko, personal communication) was used in the refinement of heavy-atom parameters. The 3.3 Å MIRAS map revealed clear electron density for a long α -helix corresponding to the heavy chain and most of the helices for the light chains. Solvent flattening and phase extension³⁰ to 3 Å resolution improved the connectivity of the electron density, especially for the light chains. An α -helical polyaniline model was constructed (using the program O; ref. 31) that included 189 of the ~360 residues in the regulatory domain, 49 of which were from the heavy chain. Phase combination of MIRAS data and the model using the program COMBINE³² improved the connectivities of the helices, and revealed the general fold of both light chains. Two calmodulin-like structures were then built for the RLC and ELC using the program XCHAIN³². Coordinates for CaM are from the Brookhaven Protein Data Bank, entry code 3CLN. The fitting of the amino-acid sequence for the ELC was aided by the mercury atom attached to each of the three sulphhydryl groups of this light chain in the ethyl mercuric phosphate (EMP) derivative. The models were further improved by several rounds of model building, refinement (using X-PLOR³³) and phase combination, which allowed an unambiguous sequence assignment for all three polypeptide chains. Two bound metal ions (Ca^{2+} in one site and Mg^{2+} and/or Ca^{2+} in the other) were then identified from difference Fourier and omit maps. Several rounds of manual rebuilding and refinement were then carried out to improve the stereochemistry. Finally, restrained individual isotropic B factors were refined. Atomic coordinates have been deposited in the Brookhaven Protein Data Bank, accession number ISCM. Abbreviations: PIP, di- μ -iodobis(ethylenediamine) diplatinum(II) nitrate; EMP, ethyl mercury phosphate; Pb, lead acetate; U, uranyl acetate.

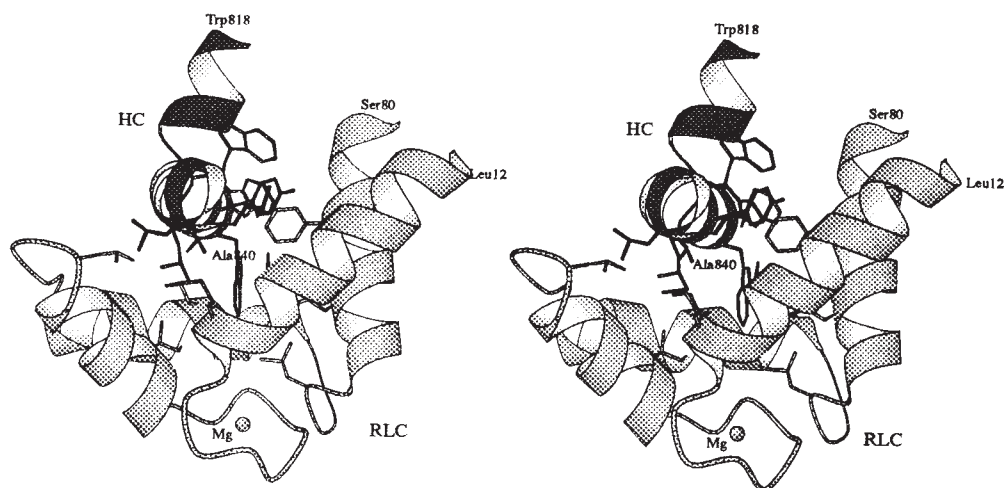
* $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity from multiple observations of the same reflection as well as its symmetry-related ones.

† The phasing power is given by r.m.s. ($|F_H|/E$), where $|F_H|$ is the heavy-atom structure factor amplitude and E is the residual lack of closure.

‡ The mean fractional isomorphous difference, $R_{\text{iso}} = \sum ||F_{\text{PH}}| - |F_P|| / \sum |F_P|$, where $|F_P|$ is the protein structure factor amplitude and $|F_{\text{PH}}|$ is the heavy-atom derivative structure factor amplitude.

§ The crystallographic R factor is given by $\sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$.

FIG. 4 Stereo drawing using Molscript³⁶ of the N-terminal lobe of the RLC and the C-terminal region of the heavy chain. The hydrophobic and aromatic residues from the heavy chain are shown interacting with similar residues from the RLC. This is the 'open' or 'gripping' state of the N-terminal lobe (see text and ref. 19).



three residues, xxY or xxW. Note that additional features in the heavy-chain sequence have not yet been identified that determine light chain versus CaM binding.

Comparison of the regulatory domain and CaM-peptide structures also strengthens the view that the light chains themselves (as well as CaM) are somewhat flexible¹⁴. Because of their arrangement in tandem, the conformations of the light chains appear to be affected by their interactions with each other as well as with the target heavy chain. The structure of the heavy chain, in turn, may also be affected by the binding, as shown by its bending near the region where the two light chains interact (see later).

RLC divalent-cation-binding site

The divalent-cation-binding site in the RLC is clearly identified by additional electron density in the region of the first loop in the N-terminal domain (Fig. 4). We do not yet know whether this site is occupied by Mg^{2+} and/or Ca^{2+} , because both cations were included in the crystallization buffer. This result confirms the previous prediction based on sequence analysis indicating

that domain I contains all the critical oxygen-donating residues required for divalent-cation binding, whereas the other three potential 'EF hands' (Fig. 2b legend) have mutations and deletions that prevent metal-ion binding¹⁵. Although the overall fold of these three defective EF hands is similar to a standard EF hand¹⁶ (including the approximate 2-fold relation between domains), only the structure of domain I approaches standard, with an angle of $\sim 90^\circ$ between helices A and B.

The only unusual amino acid in the loop of domain I is Asp 39, rather than glutamate in the twelfth position. This substitution causes the loop to be more compact than is generally observed, and may well be the reason why *in vivo* this loop coordinates Mg^{2+} (ref. 17), which has a smaller coordination radius than Ca^{2+} . Sequence data reveal that aspartate is relatively infrequent at that position in Ca^{2+} -binding proteins, but is present in a number of RLCs that bind Mg^{2+} *in vivo*¹⁸. (This inference has recently been confirmed by single-site mutagenesis (F. C. Reinach *et al.*, manuscript in preparation).) Higher resolution is required to determine the precise geometry of this metal-binding site.

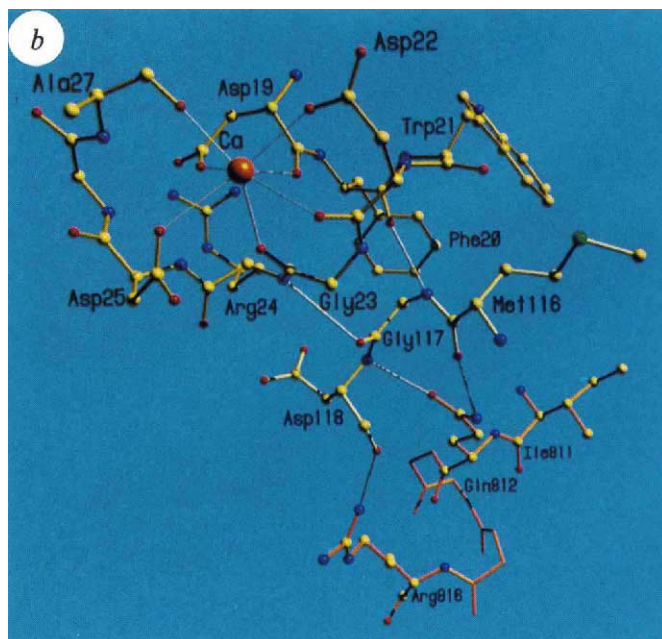
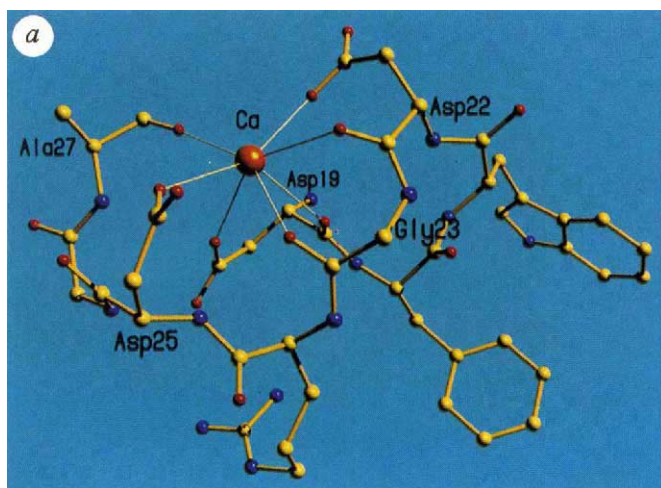


FIG. 5 a, Structure of specific Ca^{2+} -binding site in domain I of ELC showing the current model for metal coordination. A 9-residue peptide consisting of residues 19–27 loops around the Ca^{2+} ion and supplies the following ligands: the main-chain carbonyl oxygens of Asp 19, Gly 23 and Ala 27, and the carboxyl oxygens of Asp 19, Asp 22 and Asp 25. A possible seventh ligand to the carbonyl group of Asp 22 may be provided via hydrogen bonding to an intervening water molecule (latter omitted from the diagram). Higher resolution is required to establish the precise coordination geometry. b, Structure of Ca^{2+} -binding site in domain I of

the ELC showing some of the critical linkages to the RLC and heavy chain (see text). (Figures generated using a modified version of Molscript³⁶ and Rayshade (E. C. Kolb, personal communication).)

In most myosins, the RLC is more easily detached from the molecule than the ELC; in the case of scallop myosin the linkages of this subunit are especially labile. Absence of Mg^{2+} ions and an increase in temperature promote this dissociation⁴. By analogy with the structural transition pictured to occur in the two states of troponin C (ref. 19), we suggest that removal of the metal ion leads to a change of the interhelical angles in domains I and II (Fig. 4) and an internalization of the apolar residues in these domains, thereby breaking the contacts with the apolar linking residues of the heavy chain and destabilizing this part of the complex. This would be the so-called 'closed' state of the N-terminal lobe¹⁹.

ELC-specific Ca^{2+} -binding site

In contrast to the nonspecific divalent-cation-binding site on the RLC, the specific, or triggering, Ca^{2+} -binding site on ELC is unusual (Fig. 5a). Based on sequence similarities with other Ca^{2+} -binding proteins, domain III of the ELC was predicted to have a functional EF hand and to be the probable site of specific Ca^{2+} binding¹⁵. However, we have found additional electron density that we believe corresponds to a Ca^{2+} ion⁶ in domain I but not in domain III of the ELC. The loop in domain I does not have the standard pattern of EF hand, metal-liganding residues. Although there are three aspartates, their separation is unusual; moreover, there are a large number of liganding oxygen atoms from main-chain carbonyl groups. The interhelix angle between helices A and B is $\sim 135^\circ$ rather than $\sim 90^\circ$, so that the helices interlock in an antiparallel manner. Domain II of the ELC has a similar fold, but as expected from the sequence, there is no metal bound in the loop. In addition to the interhelical interactions between the two domains, stabilization is also provided by the two antiparallel β -sheets at the end of the loops.

A key feature of the specific Ca^{2+} -binding site on the ELC is that it requires the cooperative interaction of both the RLC and the heavy chain²⁰, because the ELC in isolation does not bind Ca^{2+} . Our structure shows why this is so (Fig. 5b). Domain I is the only region of the ELC that interacts with the RLC. Key contacts are two hydrogen bonds: between the main-chain carbonyl oxygen of Phe 20 of the ELC and the main-chain nitrogen of Gly 117 in domain III of the RLC, and between the main-chain nitrogen of Arg 24 of the ELC and the main-chain carbonyl oxygen of Gly 117 of the RLC. These two bonds allow very close contact between Gly 23 of the ELC and Gly 117 of the RLC. The conformation of this part of the Ca^{2+} -binding loop is therefore well stabilized, so that the carbonyl oxygen of Gly 23 can act as ligand to the Ca^{2+} ion. Stabilization of this residue is also provided by the interactions between Gln 812 of the heavy chain and Met 116 and Asp 118 on either side of the critical Gly 117 in the linker between helices F and G of the RLC. The conformation of the Ca^{2+} -binding loop is also affected by interactions between helices A and B, which make a number of contacts with the segment of the heavy chain in the second part of the IQ motif. Taken together, the structure reveals a network of critical linkages between the Ca^{2+} -binding domain and the rest of the complex.

Functional implications

The structure of the scallop regulatory domain shows how the mutual interactions of the light and heavy chains allow specific Ca^{2+} binding in a novel EF hand in domain I of the ELC. Thus, all three chains are required for regulation in the intact molecule. These results also explain findings from earlier mutagenesis studies on the light chains. Based on sequence comparisons, domain III of the ELC had been predicted to be the most probable specific Ca^{2+} -binding site¹⁵. Although this loop has the necessary disposition of potential ligand-binding residues, the structure shows that the topology of this loop does not allow metal binding. Correspondingly, mutations in this domain that would have been expected to abolish Ca^{2+} -binding are without effect⁸. More recent mutagenesis has confirmed that domain I is

the Ca^{2+} -binding site (S. Fromherz and A.G.S.-G., manuscript in preparation).

As the RLC stabilizes both the Ca^{2+} -binding loop of the ELC, as well as a segment of the heavy chain, the structure suggests how the loss of the RLC leads to the loss of both Ca^{2+} binding and regulation in the intact myosin molecule. A key residue of the RLC is Gly 117 of domain III, which is conserved in molluscan and smooth muscle myosin. Upon removal of the RLC, Ca^{2+} binding in scallop myosin can be restored only by homologous light chains from regulated myosins²¹ or by chimaeras that contain domain III of vertebrate smooth muscle²² or scallop RLCs. Moreover, mutation of the cysteine residue of vertebrate skeletal RLC to Gly converts this light chain into a functional subunit (A. Jancso and A.G.S.-G., manuscript in preparation). The structure does not, however, support conclusions based on previous crosslinking studies that the two light chains overlap extensively²³; these results may be due in part to crosslinking between two S1 heads on the same molecule.

In vertebrate smooth muscle, which is activated by phosphorylation of a serine near the N terminus of the RLC (in domain I), the region critical for regulation is also located in the C-terminal half of this light chain²⁴. Thus the sites of triggering, as well as the chemical triggers (Ca^{2+} or phosphorylation), are different in molluscan and vertebrate smooth muscle, but the interchain contacts responsible for maintaining the 'off' state may be similar.

In agreement with many previous models (reviewed in ref. 25), it has been proposed²⁶ that the light-chain-containing distal domain of S1 functions as a lever arm to amplify the conformational changes in the nucleotide- and actin-binding sites of the motor domain. The motor domain is regulated, however, by changes taking place in the distal domain, so that communication within the molecule flows in the reverse direction as well. Our results provides detailed information about some of the contacts along this pathway. Comparison of the backbone topologies of the scallop regulatory domain with the chicken S1 structure, using published stereoviews of two orientations of this region⁹, indicates that the $\sim 40^\circ$ bend in the long helix of the heavy chain appears to be similar in both molecules, although the relative dispositions of the light chains seem to be somewhat different. (We cannot, however, compare the light-chain loop regions, some of which are missing in the chicken S1 structure, and the loop in domain I of the ELC is obscured in the chicken S1 stereoviews.) Note that both molecules are in the 'on' state, which we associate with the $\sim 40^\circ$ bend, in the long helix. It is possible that a change in this bend, produced by changes in light-chain interactions (triggered by Ca^{2+} or phosphorylation), is transmitted to the motor domain. The 'top', or N-terminal, end of the long helix is very near a particularly mobile part of the motor domain, the SH1/SH2 region²⁷. There might be secondary structural changes (say coil to α -helix) upon switching that accompany very large domain rearrangements, as in some other multidomain proteins^{28,29}. In the case of the myosin molecule, the changes induced by removal of Ca^{2+} , for example, may prevent internal motor domain movements necessary for activity²⁵. The structure of the Ca^{2+} -free form of the scallop regulatory domain should provide further information about the inhibited state and the mechanism of regulation.

A full understanding of myosin control may require crystallographic studies of a regulated molecule in addition to the analysis of fragments such as the scallop regulatory domain or different S1s. (Any S1, as currently prepared, is unregulated.) But our results on the Ca^{2+} -binding switch of scallop myosin, together with the recent structure determination of chicken S1 (ref. 9), are essential steps in solving the molecular mechanism. □

Received 31 August 1993; accepted 11 January 1994.

1. Kendrick-Jones, J., Lehman, W. & Szent-Györgyi, A. G. *J. molec. Biol.* **54**, 313–326 (1970).
2. Szent-Györgyi, A. G., Szentkirályi, E. M. & Kendrick-Jones, J. *J. molec. Biol.* **74**, 179–203 (1973).

3. Wells, C. & Bagshaw, C. R. *Nature* **313**, 696–697 (1985).
4. Chantler, P. D. & Szent-Györgyi, A. G. *J. molec. Biol.* **138**, 473–492 (1980).
5. Wallimann, T. Szent-Györgyi, A. G. *Biochemistry* **20**, 1176–1187 (1981).
6. Kwon, H. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4771–4775 (1990).
7. Szentkiralyi, E. M. *J. Muscle Res. Cell Motil.* **5**, 147–164 (1984).
8. Fromherz, S. & Szent-Györgyi, A. G. *Biophys. J.* **64a**, 7 (1993).
9. Rayment, I. et al. *Science* **261**, 50–58 (1993).
10. Ikura, M. et al. *Science* **256**, 632–638 (1992).
11. Meador, W. E., Means, A. R. & Quiocho, F. A. *Science* **257**, 1251–1255 (1992).
12. Mercer, J. A., Seperack, P. K., Strobel, M. C., Copeland, N. G. & Jenkins, N. A. *Nature* **349**, 709–713 (1991).
13. Cheney, R. J. & Mooseker, M. S. *Curr. Opin. Cell Biol.* **4**, 27–35 (1992).
14. Persechini, A. & Kretsinger, R. H. *J. biol. chem.* **263**, 12175–12178 (1988).
15. Collins, J. H. *J. Muscle Res. Cell Motil.* **12**, 3–25 (1991).
16. Kretsinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248**, 3313–3326 (1973).
17. Bennett, A. G. & Bagshaw, C. R. *Biochem. J.* **233**, 173–179 (1986).
18. da Silva, A. C. R. & Reinach, F. C. *Trends biochem. Sci.* **16**, 53–57 (1991).
19. Strynadka, N. C. J. & James, M. N. *Curr. Opin. struct. Biol.* **1**, 905–914 (1991).
20. Szent-Györgyi, A. G. & Chantler, P. D. in *Myology* Vol. 2 (eds Engel, A. G. & Franzini-Armstrong, C.) 506–528 (McGraw-Hill, New York, 1994).
21. Sellers, J. R., Chantler, P. D. & Szent-Györgyi, A. G. *J. molec. Biol.* **144**, 223–245 (1980).
22. Rowe, T. & Kendrick-Jones, J. *EMBO J.* **11**, 4715–4722 (1992).
23. Hardwicke, P. M. D., Wallimann, T. & Szent-Györgyi, A. G. *Nature* **301**, 478–482 (1983).
24. Trybus, K. M. & Chatman, T. A. *J. biol. Chem.* **268**, 4412–4419 (1993).
25. Vibert, P. & Cohen, C. *J. Musc. Res. Cell Motility* **9**, 296–305 (1988).
26. Rayment, I. et al. *Science* **261**, 58–65 (1993).
27. Wells, J. A. & Yount, R. G. *Meth. Enzym.* **85**, 93–116 (1982).
28. Berchtold, H. et al. *Nature* **365**, 126–132 (1993).
29. Kjeldgaard, M., Nissen, P., Thirup, S. & Nyborg, J. *Structure* **1**, 35–50 (1993).
30. Wang, B.-C. *Meth. Enzym.* **115**, 90–112 (1985).
31. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
32. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347**, 37–44 (1990).
33. Brünger, A. T. *X-PLOR v. 3.0 Manual* (Yale Univ., New Haven, 1992).
34. Stafford, W. F., Szentkiralyi, E. M. & Szent-Györgyi, A. G. *Biochemistry* **24**, 5273–5280 (1979).
35. Funk, M. O., Nakagawa, Y., Skochdopole, J. & Kaiser, E. T. *Int. J. Pept. Prot. Res.* **13**, 296–303 (1979).
36. Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1991).
37. Sekharudu, C. Y. & Sundaralingam, M. *Prot. Sci.* **2**, 620–625 (1993).

ACKNOWLEDGEMENTS. We thank G. A. Petsko and D. Ringe for discussion and encouragement, D. Wiley and S. Harrison for access to their area detector, S. Burley and J. Kuriyan for hospitality and assistance, and P. Vibert and R. Kretsinger for discussion. This work was supported by grants from the NIH and Muscular Dystrophy Association (C.C. and A.S.G.) and by the Alexander von Humboldt Gesellschaft (I.S.). D.H.H. was supported by a grant from the NIH to G. A. Petsko and a grant from the Lucille P. Markey Charitable Trust to D. Ringe. Diffraction data were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X12-C in the National Synchrotron Light Source (supported by the US Department of Energy, Office of Health and Environmental Research and by the NSF. This article is dedicated to the memory of Eva M. Szentkiralyi.

LETTERS TO NATURE

A map of a collisionally evolving dust disk around Fomalhaut

S. Alan Stern*, Michel C. Festou† & David A. Weintraub‡

* Space Sciences Department, Southwest Research Institute, 6220 Culebra Road, San Antonio, Texas 78238, USA
 † Observatoire Midi-Pyrénées, 14 avenue E. Belin, F-31400 Toulouse, France

‡ Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA

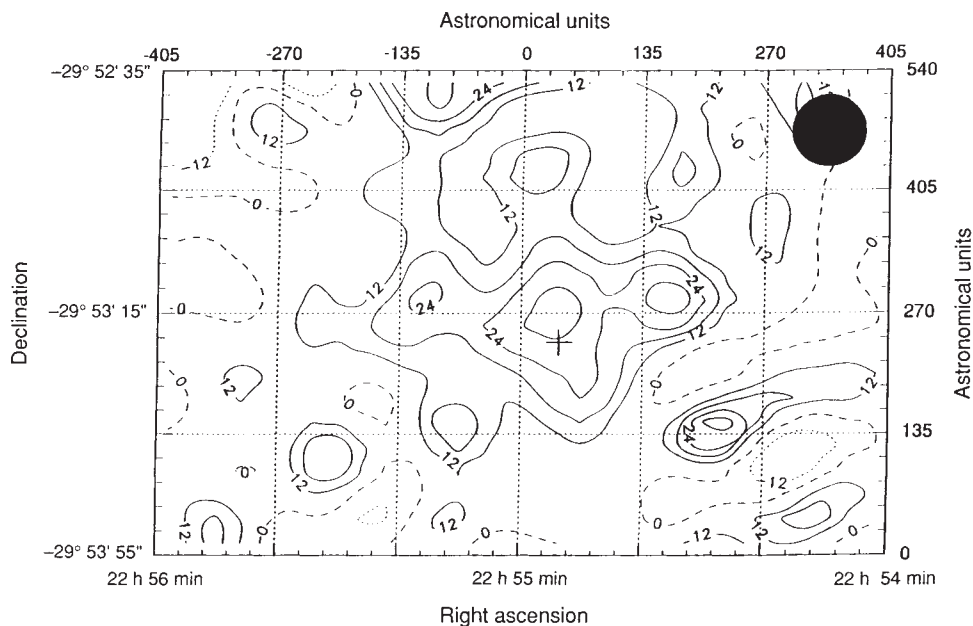
The presence of dust around normal (main-sequence) stars is a possible signature of the early stages of planet formation. Scattered light from the star β Pictoris provides evidence of a dust disk extending out to about 1,000 astronomical units¹, and the observation of far-infrared excess emission from several other main-sequence stars² suggests the presence of orbiting cold dust grains³. Because the dynamical lifetime of this dust is short, it is thought to be supplied and replenished by collisions among a population of comets or asteroids^{4–8}. Observations by Chini *et al.*^{9,10} of β Pic and several other infrared-excess stars at a wavelength of 1.3 mm have supported the idea that they are surrounded by extended dust disks. Techniques developed recently now permit imaging at these wavelengths, and we have used these techniques to obtain a map of the dust disk around the nearby, prototypical infrared-excess star Fomalhaut with a spatial resolution of 80 AU. This image provides direct confirmation that the dust is distributed in a disk-like structure, and shows that the structure extends about 200 AU from the star, farther than estimated previously^{9,10}. We estimate that a high rate of cometary and/or asteroidal collisions is required to maintain this disk.

Electromagnetic radiation from a typical main-sequence star is very nearly that of an ideal, isothermal black body. Each main-sequence stellar type has a unique temperature, characteristic of the thin, photospheric region from which it emits most of its light. From measurements at visual wavelengths, one can determine the photospheric temperature of a star and thereby directly calculate the amount of emission the star will produce as a function of wavelength. Certain stars, specifically those in the process of either forming or dying, are commonly surrounded by substantial amounts of fine-grained cold material. During the ~90%

of a star's lifetime which it spends on the main-sequence, however, the stellar environment is normally free of microscopic dust particles. The infrared-excess stars discovered by the Infrared Astronomy Satellite (IRAS) are main-sequence stars characterized by as much as an order of magnitude more emission at far-infrared wavelengths than predicted from their photospheric black-body profiles. The most likely source for this far-infrared excess radiation has convincingly been shown to be orbiting dust grains heated by the star². Because the dynamical lifetime of this dust is short, an underlying population of comets or asteroids undergoing mutual collisions is implied. As such, it is widely believed that studies of these far-infrared-excess systems provide information about the processes and timescales involved in planetary-system formation and/or evolution.

Although no successor space mission (that studies the far-infrared) to IRAS has yet flown, Chini *et al.*^{9,10}, and Zuckerman and Becklin^{7,8} have pioneered the use of the longer-wavelength submillimetre and millimetre wavelength ground-based windows to make important advances. But previous work on IRAS infrared-excess stars (such as theirs) has relied on point-flux observations made along the lines of sight to these stars. With the advent of bolometer arrays in the millimetre-wavelength range, it has now become possible to map the emission around these stars, and therefore further refine knowledge about the spatial distributions and masses of their cool dust structures. We made observations of Fomalhaut (α PsA; A3V; $D=6.7$ pc) at the Institut de Radio Astronomie Millimétrique (IRAM) telescope at Pico Veleta, Spain on 19–20 February 1993 UT, at an effective wavelength of 1.3 mm, and in excellent, very dry weather. We used the Max-Planck-Institut für Radioastronomie seven-channel bolometer¹¹, the beams of which are arranged in a hexagon surrounding a central channel. At 1.3 mm, each beam operates with a diffraction-limited half power beam width (HPBW) of 12 arcsec, corresponding to 80 AU at Fomalhaut. The centre-to-centre separation of the bolometer beams is 22 arcsec (150 AU). Because most of the observed flux in any astronomical millimetre-wave bolometer measurement is thermal emission from Earth's 100 K atmosphere, we adopted the standard procedure of using a chopping secondary mirror to remove the telluric sky signal. The secondary mirror was set to produce a chop throw of 45 arcsec at a chop frequency of 2.0 Hz. For our flux calibrator, Uranus, we adopted a 1.3-mm brightness temperature of 97.5 K (refs 12, 13). Observations of Uranus were used to calculate a system noise equivalent flux density of 80 mJy Hz^{-1/2} per channel, in good agreement with that quoted (90 mJy Hz^{-1/2} per channel) in the nominal observing performance specifications for the array.

FIG. 1 A co-added, 120×80 arcsec, 1.3-mm-wavelength map of the region around Fomalhaut made at IRAM; north is at the top, east to the left. The edges of the larger, 3×2 arcmin maps are not displayed because they were not uniformly sampled, as the central region is (see text). Fomalhaut's 1993 position is indicated by the small cross at the centre. The large black disc at the upper right represents the IRAM beam size at a wavelength of 1.3 mm (12 arcsec HPBW). There is no significant signal in the map below -1σ (-12 mJy); thus the noise level is well illustrated with the $\pm 1\sigma$ ($+12$ mJy, solid line; -12 mJy, dashed line) contours. Higher contours are drawn at 0.5σ intervals (18, 24 and 30 mJy). The 0-mJy contour is shown with a dashed line to illustrate the flatness of the subtracted baseline. The three prominent peaks near the map centre are separated by intervals of ~ 22 arcsec, which is close to (but not the same as) the spacing of the bolometer detectors; further, these three peaks do not form the hexagonal pattern of the detectors, and are not in the bolometer scan direction. We conclude they are real. The map was made in azimuth/elevation coordinates over 2.7 h, during which the airmass of Fomalhaut ranged between 2.5 and 3.1. Proper pointing was verified between maps. The sky transmission was measured three times using a standard skydip procedure. This revealed an essentially constant 1.3-mm zenith optical depth of $\tau_{1.3} = 0.13 \pm 0.01$ during the observations. The instrument was operated at highest gain for maximum sensitivity. The three maps of the Fomalhaut area were individually cleaned,



reduced to right ascension/declination coordinates, and converted to calibrated flux using the IRAM NOD2 software package^{23,24}. Single-pixel data glitches (or 'spikes') above the 3σ level were removed. The three co-registered maps were co-added and a fitted background plane was removed. This slightly tilted background plane was constrained not to remove flux at the location of Fomalhaut. The prominent feature at the centre of the map is the extended emission feature resulting from Fomalhaut's extended dust ensemble. Minor peaks in the map with lower fluxes and much smaller solid angles are due to either statistical noise and/or map edge effects.

Accurate pointing for each observation of Fomalhaut was ensured by offsetting to the position of the star after using a gaussian fit routine to centre and focus on the nearby (1° N) source 2255-282, the position of which has been well-established (G. Sandell, manuscript in preparation). Routine pointing checks revealed reliable performance with <3 arcsec (two-axis r.m.s.) pointing errors. Three, 3×2 arcmin maps were obtained by raster scanning the array over the field around Fomalhaut at an angular velocity of 4 arcsec s^{-1} , with cross-scan direction steps of one-half beam (6 arcsec). An unsmoothed, S/N-weighted co-addition of our three maps is shown in Fig. 1. (S/N is the signal-to-noise ratio). The 120×80 arcsec area shown in Fig. 1 is the fully-sampled portion of the full (180×120) arcsec map areas; partial sampling near the edges occurs because only some of the seven beams are on the map area along its edges.

Figure 1 reveals a roughly elliptical emission region centred on the 1993 position of Fomalhaut. The peak 1.3-mm emission detected toward Fomalhaut is 32 ± 12 mJy, along the line of sight to the star. The 1σ error level of 12 mJy per beam in our data (taken at high air mass but in optimum weather) compares well with the predicted sensitivity of 8 mJy. The total flux detected above the 18-mJy contour surrounding Fomalhaut is 305 ± 80 mJy (a 3.8σ detection of the extended emission). Owing to its proper motion, Fomalhaut has moved 16.8 arcsec since 1950.0. The centre of the 32-mJy isophote surrounding the peak emission corresponds (within 0.5 arcsec in right ascension and 6 arcsec in declination) to the proper-motion-corrected position of Fomalhaut. At the S/N of our data, this one-half beamwidth offset is not statistically significant. The close correspondence between the emission peak in the map and the position of the star is strong evidence that we detected radiation from dust around Fomalhaut, rather than from a randomly superimposed background source.

It is important to point out that this extended emission is not due to Fomalhaut's photosphere. This is because the flux from Fomalhaut's 8,800 K photosphere is a point source 10^4 times smaller than the IRAM beamsize, with a predicted flux of <1 mJy, and is not detectable with present-day millimetre-wavelength-techniques. As such, we conclude that the millimetre-emission region mapped around Fomalhaut is due to thermal emission from dust surrounding this main-sequence infrared-excess star. With this detection, this dust assemblage surrounding Fomalhaut becomes an important analogue to that surrounding the well-known IRAS infrared-excess star β Pictoris (ASV, $D = 16.7$ pc), which was mapped in scattered light almost a decade ago¹. Our 1.3-mm-wavelength map confirms and significantly extends earlier results^{5,7-10,14,15} which inferred (but did not directly detect) extended emission around Fomalhaut. The 18-mJy contour in Fig. 1 suggests that the emission source probably extends at least 190 AU from Fomalhaut in the east and west directions, but appears to be unresolved in the north-south direction. On the sky, the detected 18-mJy emission contour subtends 1 arcmin in the east-west direction. Although the S/N per beam of this detection is too low to make strong quantitative statements about the geometry and total flux density of the extended source, future observations with more integration time should make this possible. More-sensitive observations should also reveal whether the disk is more extended at lower flux levels than we can reliably measure in this first-detection.

As noted above, all past measurements of submillimetre- and millimetre-wavelength emission along the line of sight to Fomalhaut have been made with single-channel bolometers making point-observations. The net flux measured by that technique is the difference between the flux measured along the line of sight and the flux measured at the chop position, which is implicitly assumed to be zero. With this technique Chini *et al.*^{9,10} obtained

7.3 ± 2.2 mJy at a wavelength of 1.3 mm with a 12-arcsec HPBW and a 30-arcsec chop throw at IRAM. The same workers later detected three times as much flux, 21 ± 2.5 mJy, at 1.3 mm with a 24-arcsec HPBW and a 70-arcsec chop throw at the 15-m Swedish-ESO Submillimetre Telescope (SEST) observatory.

Our map reveals a flux along the central line of sight to Fomalhaut that appears to be higher than, but not inconsistent with, the extended emission region detected in past non-mapping observations^{7–10}. The additional flux we detected is present outside the central region. The extended nature of the Fomalhaut emission apparently revealed by the 1.3-mm map suggests that some of the point-observations of Fomalhaut made in the past may have ‘chopped away’ and/or missed flux outside their line-of-sight beams. This hypothesis is supported by the fact that Chini *et al.* found a larger flux¹⁰ in their larger beam experiment at SEST, which is consistent with their IRAM measurements only if Fomalhaut is extended at a wavelength of 1.3 mm compared to the 12 arcsec HPBW of the IRAM measurement. Given this finding, we suggest that disk properties of Fomalhaut and other infrared-excess stars studied by point-bolometer work at submillimetre–millimetre wavelengths require re-evaluation using mapping techniques. This finding has recently been described in detail¹⁶.

To estimate the total dust mass detected in the disk around Fomalhaut revealed in Fig. 1, we assume the dust grain opacity is described by a power law in frequency¹⁷

$$\kappa_\nu = \kappa_0 \left(\frac{\nu}{\nu_0} \right)^\beta$$

where κ_0 is the value of the dust grain opacity at a reference frequency ν_0 . Next, following Sandell and Weintraub¹⁶, we use eqn (6) of Hildebrand¹⁷ to write an expression for the dust mass, M_{dust} :

$$M_{\text{dust}} = 3.76 \times 10^{20} \left(\frac{1,200}{\nu} \right)^{3+\beta} \times F_\nu (e^{0.048\nu/T_{\text{dust}}} - 1) D^2 \text{ grams}$$

where ν is the frequency in GHz, F_ν is the flux in mJy, T_{dust} is the dust temperature in K, the distance D is in pc, and we adopted $\kappa_0 = 10.0 \text{ cm}^2 \text{ g}^{-1}$ at $\nu_0 = 1,200$ GHz (that is, 250 μm). This expression assumes that the emission from dust at 1.3 mm wavelength is optically thin, but does not assume that the spectrum is on the Rayleigh–Jeans tail. The optically thin assumption is strongly supported by previous IRAS and millimetre wavelength studies of Fomalhaut³.

For an integrated flux density of 305 mJy and a typical dust temperature of 43 K (expected 100 AU from Fomalhaut; the stellar luminosity $L_* = 6.7 L_\odot$, where $L_\odot$ is the stellar luminosity), we estimate a dust mass of $M_{\text{dust}} = 2.3 \times 10^{26} \times 5.2^\beta \text{ g}$. We use $\beta = 0$ and $\beta = 2$ to obtain minimum and maximum estimates of M_{dust} . This gives the range $4.4 \times 10^{26} < M_{\text{dust}} < 1.2 \times 10^{28} \text{ g}$, or

$\sim 0.07\text{--}2 M_{\text{earth}}$. Variations in the dust temperature between 30 and 100 K (as appropriate for distances of 20 to 200 AU) do not change the results by more than a factor of two. These mass estimates indicate that the mass of Fomalhaut dust detected at a wavelength of 1.3 mm is at least 10^5 times more than the interplanetary dust cloud of the Solar System¹⁹.

Now consider the timescales for loss of dust from the Fomalhaut system. The dominant dynamical loss mechanism beyond ~ 60 AU is Poynting–Robertson (P–R) drag. P–R drag limits the lifetime of low-eccentricity grains to $\sim 1 \times 10^6 R_{100}^2 \rho_{2.0} \text{ yr}$, where R_{100} is the astrocentric semi-major axis in units of 100 AU, r is the grain radius in micrometres, and $\rho_{2.0}$ is the grain density in units of 2.0 g cm^{-3} . For the estimated stellar age of Fomalhaut, $2 \pm 0.7 \times 10^8 \text{ yr}$ (ref. 20), the P–R lifetime of 30- μm grains 100 AU from Fomalhaut is 0.15 times the present age of Fomalhaut. Smaller grains are removed even faster. Even at 200 AU, the P–R lifetime for 30- μm grains against P–R drag is less than the estimated age of the star. For millimetre-sized particles, the P–R transport lifetime can approach the age of Fomalhaut at 200 AU. But using a robust ‘particle in a box’ model^{6,21,22}, it is possible to show that collisions between millimetre-sized grains in the system will remove these particles on orbits with eccentricities exceeding 1% on timescales of $10^6\text{--}10^7 \text{ yr}$, even at 200 AU. The short grain lifetimes indicated by these simple calculations provide strong circumstantial evidence for recent dust production around Fomalhaut, at least for the region inside 200 AU of interest here.

Conservatively taking $5 \times 10^7 \text{ yr}$ as a relevant dust-loss timescale, the steady-state production rate of dust needed to supply the mass of the observed Fomalhaut disk implies a total loss of 2×10^{27} to $5 \times 10^{28} \text{ g}$ ($0.3\text{--}8 M_{\text{earth}}$) over the age of the star. As 1.3-mm observations are not sensitive to either the mass locked up in large bodies (which have low surface-area to mass ratios) or in micrometre-sized grains (which are inefficient emitters at millimetre wavelengths), these dust loss and total-disk-mass estimates represent lower limits.

Because the lifetime of the observed dust is short compared to the age of Fomalhaut, it is natural to invoke an embedded population of planetesimals undergoing collisions as a mechanism to produce the observed dust. If we assume a typical instantaneous dust mass of $\sim 10^{27} \text{ g}$ and a dynamical lifetime of $5 \times 10^7 \text{ yr}$, a dust production rate of $2 \times 10^{19} \text{ g yr}^{-1}$ is implied. This is equivalent to 25–100 comet Halley masses (each $\sim 1\text{--}4 \times 10^{17} \text{ g}$) being ground up each year. This rate of collisions is $10^2\text{--}10^3$ times the rate of cometary collisions occurring in the present-day Kuiper disk surrounding the Sun^{21,22}. Recognizing the high rate of collisions implied by the dust resupply rates required at Fomalhaut, we suggest that its disk is still rapidly evolving from its formation state to a less-massive remnant disk (like our Kuiper comet disk²²). Whether planets have already formed, or are now accumulating from the colliding bodies around Fomalhaut, is an exciting area for future investigation. $\square$

Received 9 December 1993; accepted 7 February 1994.

- Smith, B. A. & Terrile, R. J. *Science* **226**, 1421–1424 (1984).
- Aumann, H. H. *et al. Astrophys. J.* **278**, L23–L27 (1984).
- Backman, D. E. & Paresce, F. in *Protostars and Planets III* (eds Levy, E. & Lunine, J. I.) 1253–1304. (Univ. of Arizona Press, Tucson, 1993).
- Weissman, P. R. *Science* **224**, 987–989 (1984).
- Harper, D. A., Lowenstein, R. F. & Davidson, D. A. *Astrophys. J.* **285**, 808–820 (1984).
- Stern, S. A. *Icarus* **84**, 447–466 (1990).
- Zuckerman, B. & Becklin, E. E. in *Submillimetre Astronomy* (eds Watt, G. D. & Webster, A. S.) 147–153 (Kluwer, Dordrecht, 1993).
- Zuckerman, B. & Becklin, E. E., *Astrophys. J.* **414**, 793–802 (1993).
- Chini, R., Krugel, E. & Kreysa, E. *Astr. Astrophys.* **227**, L5–L11 (1990).
- Chini, R., Krugel, E., Shustov, B., Tutokov, A. & Kreysa, E. *Astr. Astrophys.* **252**, 220–228 (1991).
- Kreysa, E., Lemke, R., Haslam, C. G. T. & Sievers, A. W. *Astr. Astrophys.* (in the press).
- Griffin, M. J. *et al. Icarus* **65**, 244–258 (1986).
- Orton, G. S. *et al. Icarus* **67**, 289–301 (1986).
- Gillet, F. C. in *Light on Dark Matter* (ed. Israel, F. P.) 61–69 (Reidel, Dordrecht, 1986).
- Aumann, H. H. in *The Infrared Spectra of Stars* (Pergamon, New York, 1990).

16. Weintraub, D. A. & Stern, S. A. *Astr. J.* (submitted).

17. Hildebrand, R. H. Q. *J. R. astr. Soc.* **24**, 267–282 (1983).

18. Sandell, G. & Weintraub, D. A. *Astr. Astrophys.* (submitted).

19. Sykes, M. V. *et al. in Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 336–368 (Univ. of Arizona Press, Tucson, 1989).

20. Green, E. M., Demarque, P. & King, C. R. *The Revised Yale Isochrones and Luminosity Functions* (Yale Univ. Obs., New Haven, 1987).

21. Stern, S. A. & Stewart, G. R. *Bull. Am. Astr. Soc.* **26**, 1064 (1993).

22. Stern, S. A. & Stewart, G. R. *Icarus* (submitted).

23. Haslam, C. G. T. *Astr. Astrophys. Suppl. Ser.* **15**, 333–350 (1974).

24. Salter, C. J. *IRAM Preprint No. 84* (Institut de Radio Astronomie Millimétrique, Grenoble, 1986).

ACKNOWLEDGEMENTS. We thank the staff of the IRAM observatory for their assistance. We particularly thank A. Sievers for assistance with the IRAM NOD2 bolometer image-recovery package. D. Backman, J. Bally and M. Duncan provided comments on the manuscript. G. Sandell provided us with a list of millimetre pointing sources in advance of publication. This work was supported by NASA, the NRAO (S.A.S.) and IRAM (for M.F.C.).

Generation of defects in superfluid ^4He as an analogue of the formation of cosmic strings

P. C. Hendry*, N. S. Lawson*, R. A. M. Lee*,
P. V. E. McClintock* & C. D. H. Williams†

* School of Physics and Materials, Lancaster University,
Lancaster LA1 4YB, UK

† Department of Physics, University of Exeter, Exeter EX4 4QL, UK

ALTHOUGH the birth of the Universe is inaccessible to experimental study, aspects of cosmological theories can nonetheless be explored in the laboratory. Tiny inhomogeneities in the mix of particles and radiation produced in the Big Bang grew into the clusters of galaxies that we see today, but how those inhomogeneities arose and grew is still unclear. Cosmologies based on grand unified theories suggest that a symmetry-breaking phase transition occurred via the Higgs mechanism about 10^{-34} s after the Big Bang as the Universe cooled through a critical temperature of 10^{27} K. It has been proposed by Kibble¹ that this transition may have generated defects in the geometry of space-time (such as cosmic strings), which provided the inhomogeneities on which galaxies subsequently condensed. Zurek^{2,3} has suggested that it might be possible to model this cosmological phase transition by a laboratory analogue, the superfluid transition of liquid ^4He induced by fast adiabatic expansion through the critical density. Here we report the results of such an experiment. We observe copious production of quantized vortices⁵, the superfluid analogue of cosmic strings. These results support Kibble's contention that such defects were available in the early Universe to seed galaxy formation.

The analogy^{2,4} between liquid helium and the early Universe arises because they both undergo phase transitions that can be considered as being of second order, describable in terms of Ginzburg–Landau theory⁶. In each case, the potential contribution to the free energy density can be written as

$$V = \alpha |\psi|^2 + \frac{1}{2} \beta |\psi|^4$$

where the temperature-dependent parameter α is positive above the transition and negative below it, and β is a constant. For liquid helium, the order parameter (Bose condensate wave function) ψ is given by a solution of the Ginzburg–Pitaevskii (GP) equation⁷, and there is a very similar equation⁸ to describe the gauge fields ψ in the cosmological situation. In both cases, the average, $\langle \psi \rangle$, is zero at high temperatures but as the system cools, a symmetry-breaking phase transition occurs at a critical temperature ($\sim 10^{27}$ K for the early Universe, ~ 2 K for liquid ^4He) below which $\langle \psi \rangle$ assumes finite values. Above the transition temperature, liquid ^4He exists in its normal (non-superfluid) phase, and behaves as a conventional liquid; below the transition, the coherence of the order parameter gives rise to a range of exotic properties such as frictionless flow, referred to collectively as superfluidity, in which the liquid behaves as though it were an interpenetrating mixture of two distinct components—a normal fluid and a superfluid (with zero viscosity). The early Universe above its transition temperature is believed to have existed in a symmetric state, with a so-called false vacuum characterized by unified fundamental forces and the negative pressure that provides the basis for inflationary cosmologies; below the transition, the symmetry is broken, corresponding to the true vacuum and the distinction between fundamental forces (electromagnetic, gravitational, strong and weak nuclear) that are familiar features of the Universe today.

As the system passes through the phase transition, an event horizon limits the size of causally connected regions of the nascent new phase (superfluid/true vacuum), giving rise to topological defects where the regions meet. Defect creation of this kind is generic to gauge theories with an elementary Higgs boson, for example the Standard Model and grand unified theories (GUTs). On the other hand, for theories with a composite Higgs boson, for example, Technicolour, the cosmology has not yet been studied in detail. A defect that is a feature of many cosmological models, and which appears to have the appropriate properties for nucleating galaxies, is the cosmic string—a thin tube of false vacuum. The solutions of the GP equation for superfluid ^4He , corresponding to cosmic strings, are quantized vortices⁵. Zurek's suggestion^{2,3} was that the postulated production mechanism of cosmic strings in the cosmological phase transition might be demonstrated and studied in the laboratory through the production of their superfluid analogue, quantized vortices, at the λ -transition from the normal to the superfluid state.

As a cosmological model, liquid ^4He suffers from the disadvantages that low temperatures are required and that the generated defects cannot be observed directly. On the other hand, liquid ^4He also offers important advantages. In particular, compared to liquid crystals^{9,10}, its complex scalar order parameter ψ corresponds more closely to the most commonly considered type of cosmological field, and its extremely low viscosity makes it a more plausible model of the vacuum. Furthermore, the liquid ^4He used in our experiments has been prepared in a state of extraordinarily high (arguably absolute) purity¹¹, so that one can be entirely confident that any defect formation that occurs is intrinsic to the liquid itself, and is not associated with nucleation on impurities.

Zurek argues^{2,3} that the density L of defects produced will depend on the time τ_Q taken to transverse the transition region, with

$$L = k/d^2$$

where k is the quantum of circulation in the case of ^4He , and the characteristic cell size d (in metres) of correlated regions is given in Ginzburg–Landau theory by

$$d \approx 5.6 \times 10^{-10} (\tau_Q/\tau_0)^{1/4}$$

and $\tau_0 \approx 0.85 \times 10^{-11}$ s determines the characteristic timescale of the transition. Although defects will always be generated in a passage through the transition at any finite speed, production of a large defect density requires that the transition should occur rapidly. The singularity (logarithmic infinity) in the specific heat¹² precludes the cooling of a volume of liquid helium rapidly through the λ -transition but, as Zurek pointed out, the liquid can instead be expanded through the transition at a speed limited, ultimately, by the velocity of first sound (that is, ordinary sound, a pressure–density wave in which the normal and superfluid components oscillate in phase with each other). The speed at which

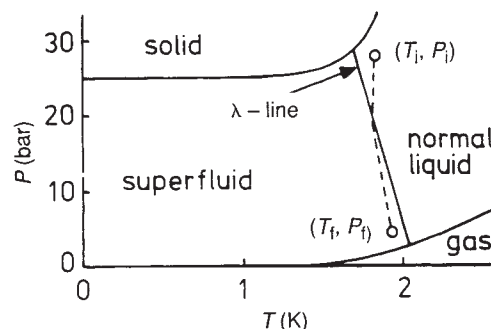


FIG. 1 Sketch of expansion trajectory (dashed) through the λ -transition on the ^4He pressure–temperature (P – T) phase diagram, from initial values (T_i, P_i) to final values (T_f, P_f) .

changes in the order parameter ψ can propagate is limited by the very much smaller velocity of second sound¹² (an entropy-temperature wave in which the normal and superfluid components oscillate in antiphase), which falls to zero at the transition itself. The phase diagram of ^4He , and a suitable expansion trajectory, are sketched in Fig. 1 (where a change in sign of the gradient of the approximately isentropic trajectory is anticipated because of the change in sign of the coefficient of expansion at the λ -transition¹²). An expansion experiment of this kind was first attempted at Los Alamos (S.-S. Shiah and J. K. Hoffer, unpublished observations), but with inconclusive results.

We have tested these ideas by expanding a small cell containing $\sim 10^{-3}$ kg of isotopically pure ^4He through the λ -transition. The cell, the body of which consists of a bronze bellows, is situated within an evacuated enclosure surrounded by a bath of He II at ~ 2 K. The temperature of the ^4He sample can be raised or lowered by means of a heater or a breakable thermal link to the bath, respectively; the volume of the sample can be adjusted by compressing or expanding the bellows by means of a mechanical linkage to the top of the cryostat. The presence of vortices produced in the expansion is detected by their attenuation of second sound^{5,12} propagated across the ~ 4 -mm space between a gold film heater and carbon bolometer situated within the cell.

Some typical signals are shown in Fig. 2, both for a time shortly after the expansion (Fig. 2a) (at early times, the signals are relatively noisy on account of some microphony of the low-level electronics, which are sensitive to the mechanical shock of the expansion) and also at a time near the end of the measuring period (Fig. 2b). Despite the noise on the early signals, it is clear that the second sound is strongly attenuated, implying the presence of vortices. The relative signal amplitude S , after normalization by division by the amplitude of the final signal in the series (that is, in the virtual absence of vortices), is plotted as a function of time in Fig. 3a: the rising signal amplitude can be attributed to a corresponding decay of the initial dense vortex tangle.

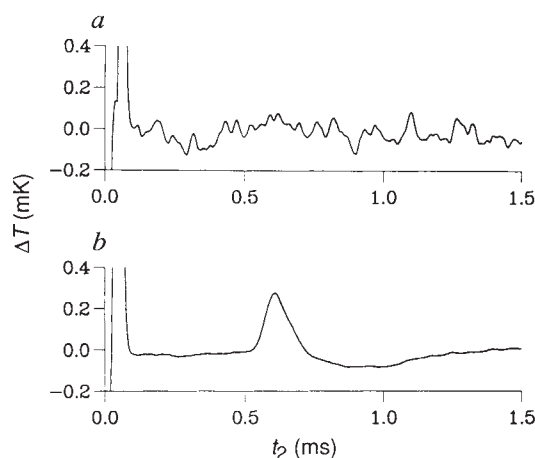


FIG. 2 Typical second-sound signals after an expansion through the λ -transition. Following the expansion, which is completed in ~ 3 ms, a sequence of ~ 100 voltage pulses is applied to the heater at 10-ms intervals, and the resultant second-sound signals arriving at the bolometer are digitized and recorded on a Nicolet 1280 data processor. The signals shown were recorded at 94 ms (a) and 1,029 ms (b) after the expansion. Each plot shows the change in temperature ΔT of the bolometer as a function of time t_2 after pulsing the heater, caused by the arrival of a pulse of second sound at $t_2 \sim 0.6$ ms. (The large excursions near $t_2 = 0$ do not represent changes in ΔT , but correspond to a direct feed-through to the detector of the voltage pulses applied to the heater.) The earlier signal, although noisy, is evidently highly attenuated compared to the later one.

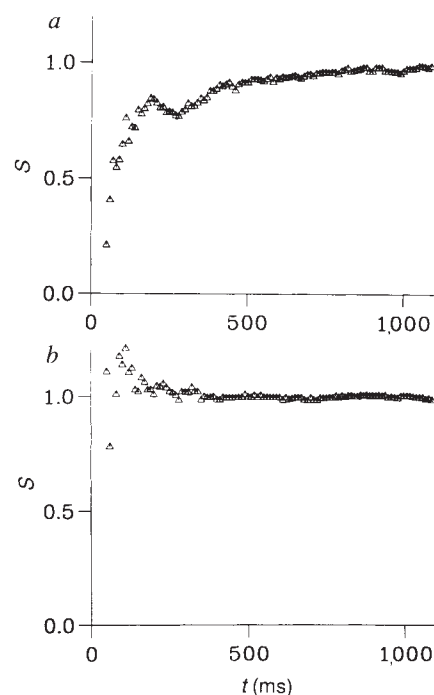


FIG. 3 Evolution of the normalized second sound signal amplitude S as a function of time t . a, Following an expansion through the λ -transition from starting temperature and pressure $T_i = 1.81$ K, $P_i = 29.6$ bar to final values of $T_f = 2.04$ K, $P_f = 6.9$ bar. b, Following a test expansion, not through the λ -transition, but lying wholly below it, with $T_i = 1.58$ K; $P_i = 23.0$ bar; $T_f = 1.74$ K; $P_f = 4.0$ bar.

In addition to the vortices produced by the Zurek mechanism^{2,4}, some vortices will also have been generated by more conventional hydrodynamical means. This is because the expansion of a real experimental cell is non-ideal, in the sense that some flow of liquid parallel to surfaces is bound to occur (for example, on account of the convolutions of the bellows, and the presence of the heater and bolometer). Such flows will have exceeded the critical velocity for vortex creation⁵, but only for a very short time so that the resultant vortex density is expected to be small. Figure 3b shows the results of a test expansion starting well below the λ -transition, in which the fluid flows will have been very similar to those for trajectories passing through the transition. There is no evident attenuation: although some vortices must certainly have been produced, they are undetectable in the present experiment. Expansions from very slightly (a few milliKelvin) below the transition also result in significant transient second sound attenuation, although weaker than that arising from expansions passing through the transition; the origin of the corresponding vortices remains unclear. They may perhaps be associated with the vortices¹³ generated by the large fluctuations of ψ that occur in thermal equilibrium just below the transition, although recent calculations (W. F. Vinen, manuscript in preparation) appear to weigh against the possibility: the resolution of this interesting question may be expected to illuminate the physical nature of the λ -transition.

It follows that the strong transient attenuation of second sound caused by expansion through the λ -transition can be taken as evidence for vortices produced through the Zurek mechanism^{2,4}. The form of the subsequent vortex decay, indicated by the growing signal in Fig. 3a, is very similar to that seen or discussed in earlier work in that there are two distinct decay rates^{14,16}, sometimes separated by a local maximum. On the assumption that the fast decay rate varies as L^{-2} , where L is the vortex line density, the initial fast rate can be used to extrapolate an estimate of the initial line density L_i immediately following

the expansion: we find that $L_i \approx 10^{11} \text{ m}^{-2}$. This is a conservative estimate in that the experimental error leads to an upper bound that is larger by several orders of magnitude, and a lower bound of $5 \times 10^{10} \text{ m}^{-2}$. Although more accurate values of L_i will be obtained when the sensitivity of the electronics to mechanical shock has been reduced, it is already clear that the density of vortices generated is very large. To put these figures into perspective, note that the production of a line density of $\sim 10^{11} \text{ m}^{-2}$ by

rotation⁵ of a vessel of liquid ^4He would require an angular velocity of $\sim 4,000 \text{ rad s}^{-1}$, an enormous value.

This copious production of vortices created, as predicted^{2,3}, by passage through the λ -transition implies, by analogy, that cosmic strings were generated in the mathematically similar second-order phase transition of the early Universe, and that they would therefore have been available to play the role in galaxy formation proposed by Kibble¹. $\square$

Received 20 December 1993; accepted 18 February 1994.

1. Kibble, T. W. B. *J. Phys.* **A9**, 1387–1398 (1976).
2. Zurek, W. H., *Nature* **317**, 505–508 (1985).
3. Zurek, W. H. *Acta phys. pol.* **B24**, 1301–1311 (1993).
4. Kibble, T. W. B. *Nature* **317**, 472 (1985).
5. Donnelly, R. J. *Quantized Vortices in Helium II* (Cambridge Univ. Press, 1991).
6. Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* 2nd edn (Hilger, Bristol, 1986).
7. Ginzburg, V. L. & Pitaevskii, L. P. *Soviet Phys. JETP* **7**, 858–861 (1958).
8. Vilenkin, A. *Phys. Rep.* **121**, 263–315 (1985).
9. Chuang, I., Turok, N. & Yurke, B. *Phys. Rev. Lett.* **66**, 2472–2475 (1991).
10. Bowick, M. J., Chander, L., Schiff, E. A. & Srivastava, A. M. *Science* **263**, 943–945 (1994).

11. Hendry, P. C. & McClintock, P. V. E. *Cryogenics* **27**, 131–138 (1987).
12. Wilks, J. *The Properties of Liquid and Solid Helium* (Clarendon Press, Oxford, 1967).
13. Williams, G. A. J. *Low Temp. Phys.* **89**, 91–100 (1992).
14. Vinen, W. F. *Proc. R. Soc. A* **242**, 493–515 (1957).
15. Schwarz, K. W. & Rozen, J. R. *Phys. Rev. Lett.* **66**, 1898–1901 (1991).
16. Smith, M. R., Donnelly, R. J., Goldenfeld, N. & Vinen, W. F. *Phys. Rev. Lett.* **71**, 2583–2586 (1993).

ACKNOWLEDGEMENTS. Discussions with D. H. Lyth, H. J. Maris, E. N. Smith, W. F. Vinen, G. A. Williams and W. H. Zurek are acknowledged. The research is supported by the UK Science and Engineering Research Council.

Generalized synthesis of periodic surfactant/inorganic composite materials

Qisheng Huo*, David I. Margolese*,
Ulrike Ciesla†, Pingyun Feng*, Thurman E. Gler*,
Peter Sieger*, Rosa Leon‡, Pierre M. Petroff‡,
Ferdi Schüth§ & Galen D. Stucky§

* Department of Chemistry, University of California, Santa Barbara, California 93106, USA

† Institut für Anorganische Chemie, Johannes-Gutenberg Universität, 55099 Mainz, Germany

‡ Center for Quantized Electronic Structures, University of California, Santa Barbara, California 93106, USA

THE recent synthesis of silica-based mesoporous materials^{1,2} by the cooperative assembly of periodic inorganic and surfactant-based structures has attracted great interest because it extends the range of molecular-sieve materials into the very-large-pore regime. If the synthetic approach can be generalized to transition-metal oxide mesostructures, the resulting nanocomposite materials might find applications in electrochromic or solid-electrolyte devices^{3,4}, as high-surface-area redox catalysts⁵ and as substrates for biochemical separations. We have proposed recently⁶ that the matching of charge density at the surfactant/inorganic interfaces governs the assembly process; such co-organization of organic and inorganic phases is thought to be a key aspect of biomineralization⁷. Here we report a generalized approach to the synthesis of periodic mesophases of metal oxides and cationic or anionic surfactants under a range of pH conditions. We suggest that the assembly process is controlled by electrostatic complementarity between the inorganic ions in solution, the charged surfactant head groups and—when these charges both have the same sign—inorganic counterions. We identify a number of different general strategies for obtaining a variety of ordered composite materials.

Four pathways to the synthesis of mesostructured surfactant-inorganic biphasic arrays are depicted in Fig. 1. Route 1 involves the direct co-condensation of anionic inorganic species with a cationic surfactant (S^+I^-); the syntheses of MCM-41 and MCM-48 (refs 1, 2) are prototypic examples. We have now found a similar route to periodic tubular non-silica structures, but involv-

ing cooperative condensation of a cationic inorganic species with an anionic surfactant (S^-I^+), (route 2) a possibility predicted in ref. 6. By contrast, routes 3 and 4 involve condensation of ionic inorganic species in the presence of similarly charged surfactant molecules. These pathways are mediated by counterions of opposite charge to that of the surfactant head group (solution species ($\text{S}^+\text{X}^-\text{I}^+$) where $\text{X}^- = \text{Cl}^-, \text{Br}^-$; or, ($\text{S}^-\text{M}^+\text{I}^-$) where $\text{M}^+ = \text{Na}^+, \text{K}^+$).

Lamellar, MCM-41 and MCM-48 periodic porous silicates (S^+I^-) are obtained under basic conditions by the self-assembly of anionic silicates and cationic surfactant molecules^{1,2,6,8}. We suggested in ref. 6 that this mechanism should work for other oxides as well, if the pH is varied to adjust the charge density of the metal oxide polyanions. Thus, we were able to synthesize lamellar tungsten (vi) oxide ($d_{100} = 31 \text{ Å}$, where d_{100} is the first X-ray diffraction line) at pH above 8, and lamellar and hexagonal tungsten (vi) oxide phases ($d_{100} = 40 \text{ Å}$) at pH 4–8 (Fig. 2) using cationic surfactants. As another example, both cubic ($Ia3d$, pH 6.5–7.0) and hexagonal (pH 6.0–6.5) antimony (v) oxide mesostructures were synthesized in lower-pH windows for the anionic inorganic phase. By choosing inorganic species (for example, antimony or tungsten oxides) that are more acidic than silicic acid, the cooperative biphasic templating can be carried out at neutral and lower pH values to obtain a match of the charge density of the oxide to that of the respective surfactant phase.

A similar approach was taken in a charge-reversed situation, where an anionic surfactant was used to direct the condensation of cationic oxide species (S^-I^+). $\text{C}_{16}\text{H}_{33}\text{SO}_3\text{H}$ was used in the synthesis of iron and lead oxides, giving a hexagonal phase ($d_{100} = 45.8 \text{ Å}$ for Pb) and different lamellar phases (for example, 38.5 Å for Pb and 41.0 Å for Fe). Lamellar phases were obtained depending on the synthesis conditions (pH 1–5), using oxides of Mg, Al, Mn, Fe, Co, Ni and Zn with $\text{C}_{12}\text{H}_{25}\text{PO}_4\text{H}_2$ as the surfactant. An interesting example is aluminium oxide with sodium dodecylbenzenesulphonate at 70°C and pH 3.5, which slowly rearranges to increase the layer spacing (28.9 to 32.1 Å in 8 hours to 9 days). It should be noted that the anionic polar head group can be a part of the inorganic framework in these materials.

Surprisingly, the formation of mesophases was possible by the cooperative assembly of cationic inorganic species with cationic surfactants. Generally lamellar, hexagonal and $Pm3n$ cubic phases (Fig. 1) were prepared by addition of tetraethyl-orthosilicate (TEOS) at room temperature to an acidic (for example, HCl) solution of surfactant. After stirring for 30 min or longer, the solid product was recovered by filtration and studied by X-ray diffraction and high-resolution transmission electron micros-

§ To whom correspondence should be addressed.

copy (Figs 2, 3). High-quality samples of the three phases can be formed in a wide range of strongly acidic conditions (1–7 M HCl or HBr). We found that HCl favours the formation of all three phases, whereas HBr favours the formation of the hexagonal and not the cubic phase. In contrast to the synthesis conditions used previously^{1,2,6} for M41S silicate mesophases, extremely acidic media, low temperatures, short synthesis times and low concentrations of surfactant were employed. Zwitterionic and lipid-like surfactants were successfully used in this synthesis. Silica polymerization proceeds in this pH range through the condensation of cationic intermediates^{9,10}.

The new cubic silica phase, first reported here (Figs 1–3), was prepared using surfactants with larger head groups (for example, alkyltrimethylammonium, cetylpyridinium) as the template. This was done to decrease the value of the surfactant packing parameter to generate maximum surface curvature^{11,12}. The X-ray diffraction data closely matches (intensities and indexing) that for the I_1 ($Pm3n$) phase of cetyltrimethylammonium chloride and cetylpyridinium chloride in formamide¹³. This phase consists of a packing of two types of discrete micellar aggregates^{13–17} in globular-like cages, with one large cage of ~ 30 Å diameter giving a clathrasil-like structure. Removal of the $C_{16}TEA^+$ (triethylammonium) template by calcination at 500 °C resulted in a reduction of the unit cell parameter from 89.9 to 78.2 Å.

The surfactant–silica mesophases exhibit different unit cell parameters when surfactants with differing alkyl chain lengths were used (Fig. 5). Moreover, the X-ray diffraction data (Fig. 5), show clearly that similar surfactants at pH < 0 produce mesostructures with similar, but slightly larger d -spacings than those obtained by basic (pH > 9) medium synthesis. For the $C_{16}TMA^+$

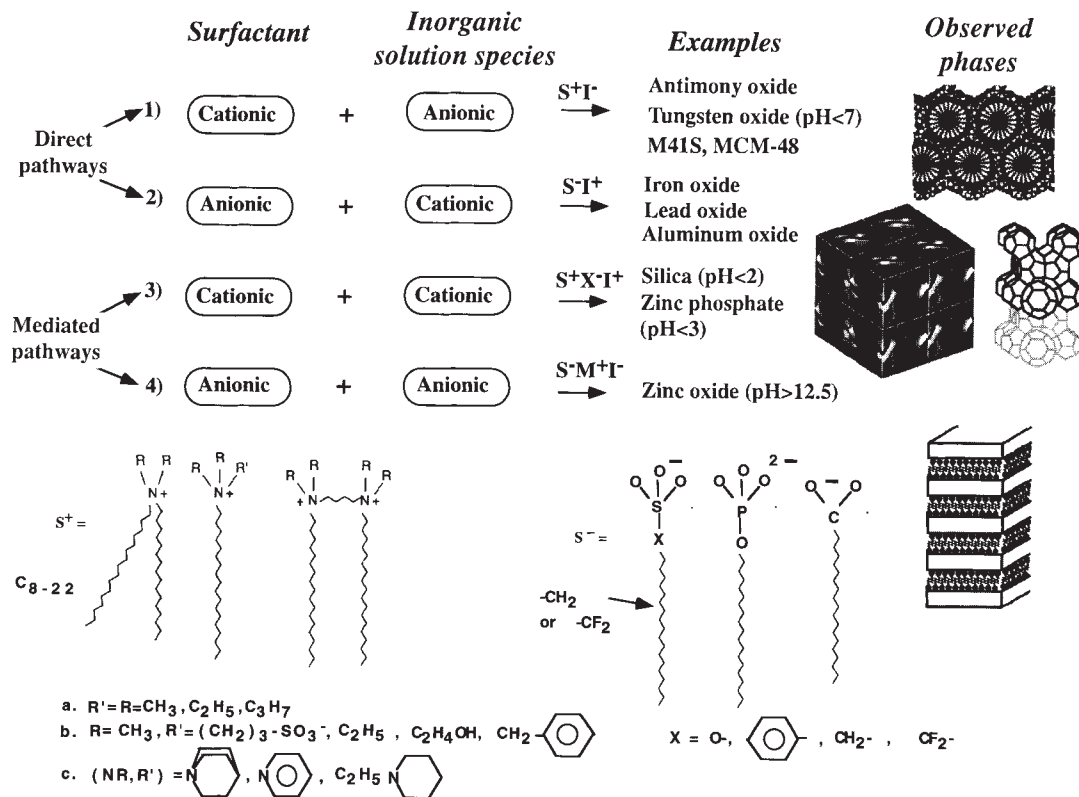
(trimethylammonium) hexagonal phase product formed in acidic media, chemical analysis gave (wt%): SiO_2 , 39.7; $C_{16}TMA^+$, 46.38; Cl, 5.75. The product surfactant/Cl ratio is therefore 1.0:1.0, a feature that distinguishes these mesostructure phases from hexagonal phase MCM-41 formed in alkaline solution which has no significant halide content using the same surfactant halide starting materials. Because in the acid meso-phase product the template cationic charge is exactly balanced by a halide anion, the template should be removable without providing an exchangeable cation, as required for the anionic frameworks of MCM-41 (ref. 18). Indeed, it was found that >85% of the template could be removed from the material as-synthesized by stirring in pure ethanol at room temperature or by overnight reflux. The removal of the template was also possible by calcination. The calcined hexagonal phases (at 500 °C) had surface areas greater than $1,000\text{ m}^2\text{ g}^{-1}$ and average pore diameters of 18.6, 26.5 and 33.2 Å for C_{14} , C_{16} and C_{18} TMA^+ , respectively, as determined by N_2 BET (Brunauer–Emmett–Teller) measurements.

^{29}Si MAS (magic angle spinning) NMR data (Fig. 4) were used to determine the degree of polymerization and the concentration of silanol, as measured by the ratio of the two major peaks (denoted Q^3 and Q^4). At long reaction times, for the hexagonal phase, Q^3/Q^4 decreases due to increased polymerization, but for a given reaction time, Q^3/Q^4 is approximately the same (0.55) as for the hexagonal material precipitated from basic medium.

As proposed for the co-condensation of oligomeric polysilicate and cationic surfactant molecules in alkaline solution (pH > 9)^{1,2,6}, we believe that the main driving force for self-assembly in strong acid media is electrostatic. We postulate that

FIG. 1 A general scheme for the self-assembly reaction of different surfactant and inorganic species.

The mesostructured phases can be synthesized by either direct pathways or mediated pathways, using different charged surfactant molecules and inorganic solution species in certain pH conditions. Some surfactants other than alkyltrimethylammonium ions are also good templates for mesoporous materials syntheses. Each surfactant can act as a template for the formation of one or more mesostructures in different reaction conditions. The topology of the composite mesostructures reflects the final geometry adopted by the organized organic array so that regardless of the inorganic framework composition, liquid-crystal morphologies have been observed and can be expected. The dimensionless effective surfactant packing parameter $g^* = V/a_0l$ determines the product structure^{11,12,25}, where V is the chain volume, l is the chain length and a_0 is the optimum head-group area of a surfactant molecule.



the interactions between cationic silica species and halide-cationic surfactant headgroups are mediated by the halide ions, which form hydrogen bonds to protonated silanols such as $\equiv\text{Si}(\text{OH}_2)^+$ (ref. 10). At high concentrations of hydrogen halide solutions, the cationic surfactant hydrophilic region is surrounded by halide ions forming an electrical double layer (S^+X^-) with the periphery negatively charged. As cooperative assembly and precipitation proceeds, the protons associated with the silica species along with associated excess halide ions are excluded until a neutral inorganic framework remains.

This model is supported by: (1) the fact that cationic silica species are present at $\text{pH} < 2$ (refs 9,10) and the fact that the solution $[\text{H}^+]$ does not change during synthesis; (2) the 1:1 surfactant to chlorine ratio in the hexagonal product; (3) the easy removal of the template with ethanol; (4) the observation

that TEOS and SiCl_4 hydrolyse and form mesophase products, whereas Cab-O-Sil (SiO_2) which does not dissociatively hydrolyse in acidic conditions, forms no mesophase products. Clearly the template mechanism for (S^+X^-) synthesis is not the same as that for (S^+I^-) synthesis.

We have applied the above approach to the synthesis of organized Zn-containing phosphate mesostructured materials. At least four lamellar $(\text{C}_n\text{TMA})^+\text{X}^-[\text{H}_2\text{ZnPO}_4]$ (X is Cl^- , Br^-) phases with different d -spacings and surfactant packing, including monolayer and bilayer geometries, can be generated by controlling composition and pH. One of them was formed from solutions containing $[\text{H}_2\text{ZnPO}_4]^+$ ($\text{pH} \sim 2.5$) (Figs 3, 5), and like the acid silica phases contains one halide ion per surfactant molecule. In this structure there is three-dimensional ordering and excellent lateral ordering of layers. Comparison with the

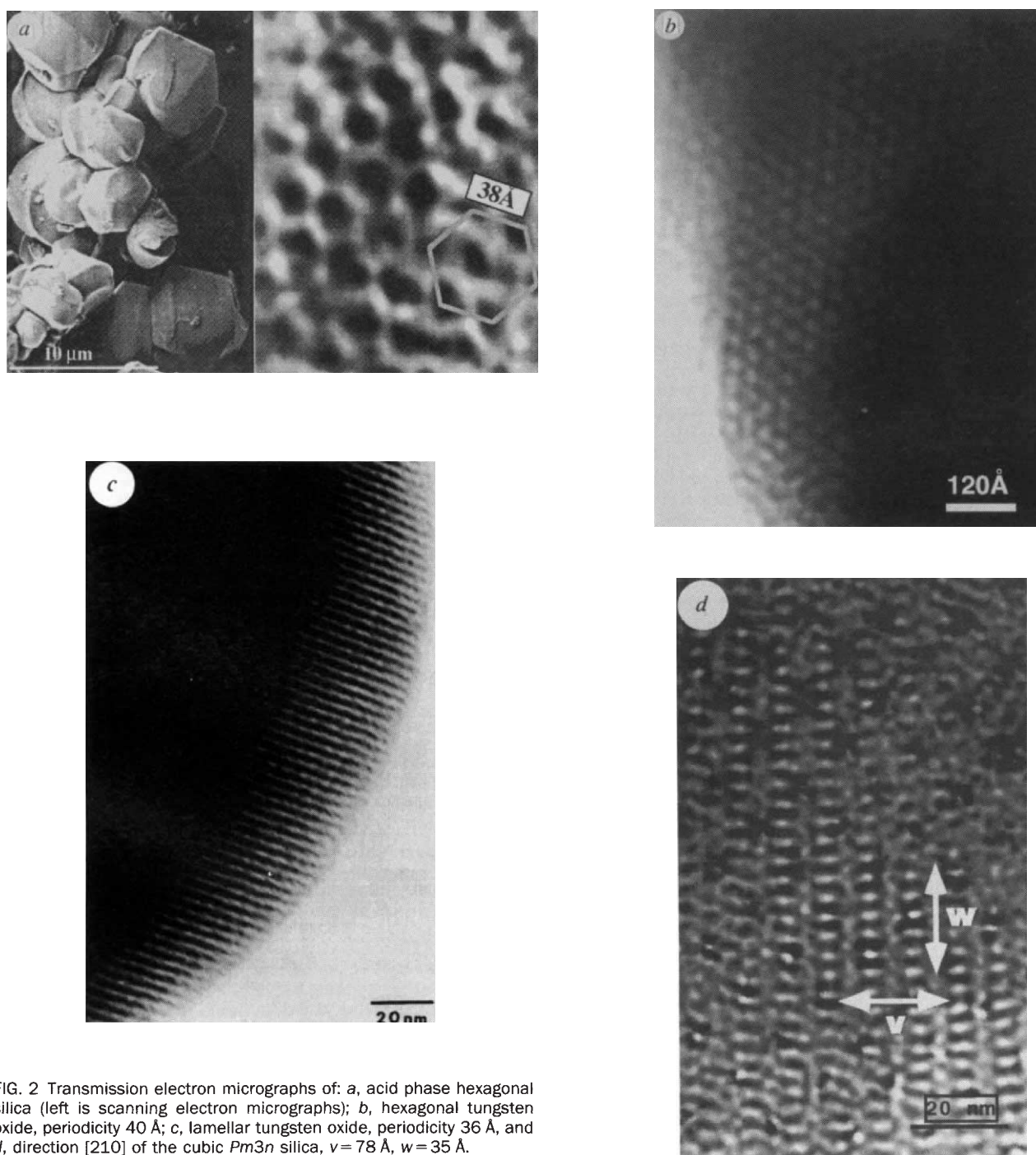


FIG. 2 Transmission electron micrographs of: a, acid phase hexagonal silica (left is scanning electron micrographs); b, hexagonal tungsten oxide, periodicity 40 Å; c, lamellar tungsten oxide, periodicity 36 Å, and d, direction $[210]$ of the cubic $Pm\bar{3}n$ silica, $v = 78$ Å, $w = 35$ Å.

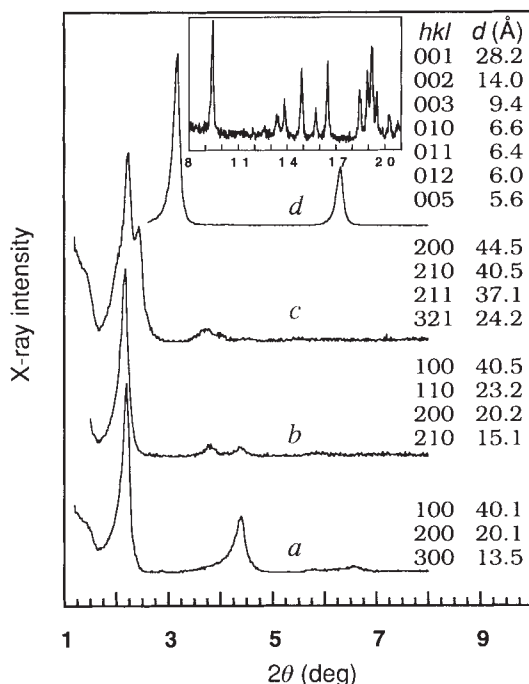


FIG. 3 Powder X-ray diffraction (XRD) patterns of silica and zinc phosphate mesostructures precipitated from acidic medium. From bottom to top, the patterns are: (a) lamellar silica, $a_0 = 39.8$ Å; (b) hexagonal silica (MCM-41), $a_0 = 47.6$ Å; (c) cubic $Pm\bar{3}n$ silica, $a_0 = 89.9$ Å; (d) lamellar zinc phosphate, with the high- 2θ region shown in the inset. These peaks (d) could be indexed in an orthorhombic cell having $a = 6.49$ Å, $b = 6.57$ Å and $c = 28.08$ Å. Typical preparation of the silica mesophases from acidic medium was carried out with the following molar ratios: (a) lamellar: 0.12 C_{20} TMA-Br:1TEOS:4.9 HCl:130 H_2O ; (b) hexagonal: 0.12 C_{16} TMA-Br:1 TEOS:9.2 HCl:130 H_2O and (c) cubic: 0.13 C_{16} TEA-Br:1 TEOS:10.4 HCl:130 H_2O . The lamellar zinc phosphate phase shown here (d) was synthesized by crystallization of a reaction mixture with molar composition of 3.0 $Zn(NO_3)_2$:3.5 H_3PO_4 :5.0 C_{16} TMA-Br:4.5 TMA-OH:1,000 H_2O at 4 °C for 18 d (pH 2) to give a final composition of $[(C_{16}H_{33})N(CH_3)_3^+Br^-] \cdot HZnPO_4$, [observed, calculated: Zn (12.4%, 12.5%); Br (15.4%, 15.2%); N (2.9%, 2.7%); and P (5.5%, 5.9%)]. By varying pH and dilution, other lamellar structures can be obtained with spacings of 63, 37.3 and 29.7 Å and surfactant chain length to lattice spacing variation which suggests a surfactant bilayer configuration for the first of these. In these lamellar phases, the residue obtained after thermal analysis to 900 °C is $Zn_2P_2O_7$.

determined structures of the layered $MH(ZnPO_4)_2$ (where M is Na, Cs) (ref. 19) strongly supports a template model similar to that for the acid synthesized silica phases, with the cationic surfactant indirectly coordinated to the $HZnPO_4$ walls by hydrogen bonding to the anion (Cl^- or Br^-).

Finally, we have explored the possibility of synthesizing periodic biphasic arrays from negatively charged surfactants and inorganic species (S^-I^-). For instance, $Zn(OH)_3^-$, $Zn(OH)_4^{2-}$ and $Zn_2(OH)_6^{2-}$ are the predominant species in a 0.1 M solution of ZnO (pH 10–14) (ref. 20). The anionic surfactant $CH_3(CH_2)_{16}COO^-M^+$ gives a lamellar zinc oxide phase ($d_{001} = 48.8$ Å, 50.3 Å; $M^+ = Na^+$, K^+ , respectively) above pH 12.5 (TMA-OH without surfactant under the same reaction conditions gives only ZnO), consistent with a mediated templating ($S^-M^+I^-$) pathway. Anionic surfactants compete with oxide/hydroxide groups for framework metal atom coordination sites and under similar conditions phosphates such as $C_{12}H_{25}OPO_3Na_{2-x}H_x$ ($x = 0-1$) appear to coordinate directly to the zinc atom, thus becoming part of the framework with no sodium ions present in the final product. In another example, at pH > 8 using an anionic aluminate species in solution lamellar ($d_{001} = 33.6$ Å) $[C_{12}H_{25}OPO_3NaH]_3 \cdot Al(OH)_3 \cdot 4H_2O$ (observed and calculated Al (2.5%, 2.5%); P (8.6%, 8.6%); C (40.2%, 39.7%); H (7.7%, 8.9%); Na (5.56%, 6.35%)) is formed in contrast to lamellar $[C_{12}H_{25}OPO_3] \cdot Al(OH)_3 \cdot 2H_2O$ (observed and calculated Al (7.9%, 8.2%); P (9.3%, 9.5%); C (44.4%, 44.2%); H (8.6%, 8.6%)) which forms at pH < 5. The former is consistent with a ($S^-M^+I^-$) solution pathway and no framework phosphate, whereas in the latter the (S^-I^+) chemistry prevails with surfactant framework participation.

The results presented here confirm the feasibility of using the synergistic interface chemistry of inorganic and organic species to create periodic mesostructures with cationic or anionic inorganic species including transition-metal oxides and cationic or anionic surfactants. A variety of surfactants including lipids and zwitterionic types were used so that this chemistry becomes directly applicable to biomimetics. Our periodic biphasic arrays are to be contrasted with the tubules of silica²¹ and iron oxide²² grown on a charged lipid template. Charge matching of the biphasic arrays can be accomplished through the four pathways shown in Fig. 1, but even these may not be exclusive as charge

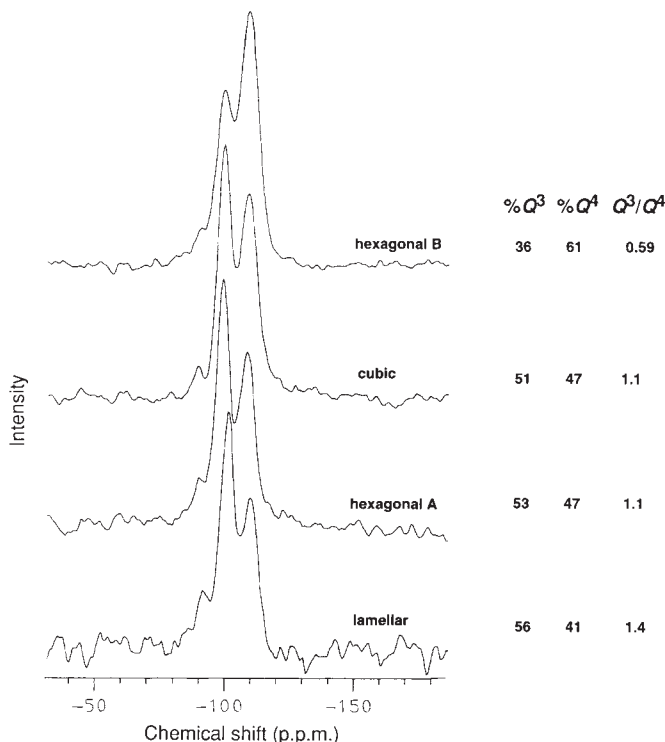


FIG. 4 ^{29}Si MAS NMR of the as-synthesized silica mesophases. For all four of the materials the two major peaks shown are Q^3 (at -101 to -102 p.p.m.) and Q^4 (at -110 to -111 p.p.m.) (ref. 26). % Q^3 measures the number of ^{29}Si of the type $(SiO)_3 \equiv Si-OH$ and % Q^4 measures the number of ^{29}Si of the type $(SiO)_4 \equiv Si-O-Si-$. The estimated error in the values of Q^3 and Q^4 are $\pm 5\%$ of their value. Hexagonal phase A was run at short reaction time (30 min) and hexagonal phase B was run under the same synthesis conditions except for long reaction time (2 h). The Q^3/Q^4 ratio measures the extent of silanol condensation indicating for material B a more condensed framework than material A. Q^3/Q^4 for the hexagonal phase and for the lamellar phase in acidic media is similar to that found for materials precipitated from basic media, which are 0.55 and 1.2 respectively. Chemical shifts were measured with respect to tetramethylsilane.

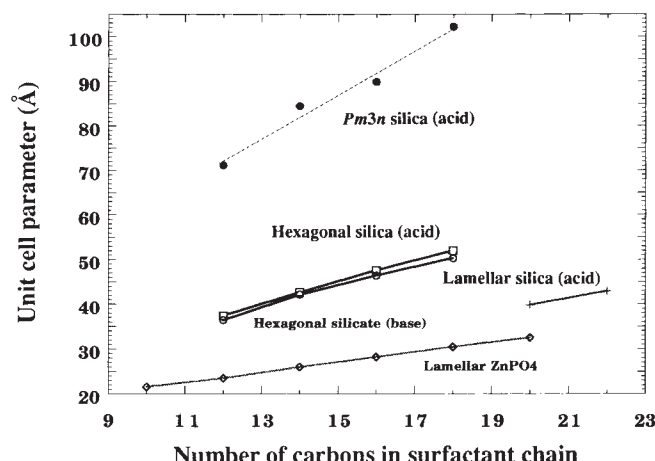


FIG. 5 The relationship between unit cell parameter and number of carbon atoms in surfactant chain. Silica (*Pm3n*) phase was synthesized by using alkyltrimethylammonium as template at room temperature for 1 h. Hexagonal silica and silicate phases were obtained in the presence of alkyltrimethylammonium at room temperature in acidic and basic media respectively. Only $C_{20-22}TMA^+$ led to the formation of lamellar silica in acidic medium at room temperature. Lamellar $ZnPO_4$ was prepared by using alkyltrimethylammonium as template from acidic medium (pH 2.5). Estimated error is 1 Å in *d*-spacing of highest peak (d_{210} for *Pm3n* phase and d_{100} for hexagonal and lamellar phases).

matching also might be accomplished through a combination of both direct and mediated linkages. Ultimately, temporal and spatial control of the relative rates of the biphasic and interface assembly processes in cooperative templating might be used to impose microscopic long-range modulation of the periodic composite array morphology in a manner directly relevant to biomaterial synthesis^{23,24}.

Received 21 September 1993; accepted 27 January 1994.

1. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
2. Beck, J. S. *et al.* *J. Am. chem. Soc.* **114**, 10834–10843 (1992).
3. Lampert, C. M. & Granqvist, C. G. in *Large-Area Chromogenics: Materials and Devices for Transmittance Control* (eds Lampert, C. M. & Granqvist, C. G.) 2–19 (SPIE Optical Engineering, Washington DC, 1990).
4. Dautremont-Smith, W. C. *Displays* **3**, 3–22 (1982).
5. Parton, R. F., Jacobs, J. M., van Ooteghem, H. & Jacobs, P. A. *Zeolites as Catalysts, Sorbents and Detergent Builders* (eds Karge, H. G. & Weitkamp, J.) 211–212 (Elsevier, Amsterdam, 1989).
6. Monnier, A. *et al.* *Science* **261**, 1299–1303 (1993).
7. Mann, S. *Nature* **365**, 499–505 (1993).
8. Inagaki, S., Fukushima, Y. & Kuroda, K. *J. chem. Soc., chem Commun.* 680–682 (1993).
9. Iler, R. K. *The Chemistry of Silica* (Wiley, New York, 1979).
10. Brinker, C. J. & Scherer, G. W. *Sol-Gel Science* 97–233 (Academic, San Diego, 1990).
11. Hyde, S. T. *Pure appl. Chem.* **64**, 1617–1622 (1992).
12. Israelachvili, J. N., Mitchell, D. J. & Ninham, B. W. *J. chem. Soc., Faraday Trans. 2* **72**, 1525–1568 (1976).
13. Auvray, X. *et al.* *Langmuir* **9**, 444–448 (1993).
14. Vargas, R., Mariani, P., Gulik, A. & Luzzati, V. *J. molec. Biol.* **225**, 137–145 (1992).
15. Charvolin, J. & Sadoc, J. F. *J. Phys. Paris* **49**, 521–526 (1988).
16. Seddon, J. M. & Templer, R. H. *Phil. Trans. R. Soc. A* **344**, 377–401 (1993).
17. Fontell, K. *Colloid Polym. Sci.* **268**, 264–285 (1990).
18. Whitehurst, D. D. US Patent No. 5143879 (1992).
19. Nenoff, T. M., Harrison, W. T. A., Gler, T. E., Galabrese, J. C. & Stucky, G. D. *J. Solid State Chem.* **107**, 285–295 (1993).
20. Baes, C. F. Jr & Mesmer, R. E. *The Hydrolysis of Cations* 287–294 (Wiley International, New York, 1976).
21. Barai, S. & Schoen, P. *Chem. Mater.* **5**, 145–147 (1993).
22. Archibald, D. D. & Mann, S. *Nature* **364**, 430–433 (1993).
23. Friedbacher, G., Hansma, P. K., Ramil, E. & Stucky, G. D. *Science* **253**, 1261–1263 (1991).
24. Morse, D. E., Carliou, M. A., Stucky, G. D. & Hansma, P. *Mat. Res. Soc. Symp. Proc.* **292**, 59–67 (1993).
25. Rosen, M. J. *Surfactants and Interfacial Phenomena* 108–142 (Wiley, New York, 1989).
26. Lippmaa, E., Magi, M., Samoson, A., Engelhardt, G. & Grimmer, A. R. *J. Am. chem. Soc.* **102**, 4889–4893 (1980).

ACKNOWLEDGEMENTS. We thank S. Mann for suggestions and comments. This work was supported by the US Office of Naval Research (G.D.S. and D.I.M.), the US NSF (G.D.S. and Q.H.), Air Products (G.D.S., T.E.G. and Q.H.), DFG (P.S., U.C. and F.S.) and the US NSF Science and Technology Center (R.L. and P.M.F.). We made use of MRL Central Facilities supported by the US NSF.

Titanium-containing mesoporous molecular sieves for catalytic oxidation of aromatic compounds

Peter T. Tanev, Malama Chibwe & Thomas J. Pinnavaia*

Department of Chemistry and Center for Fundamental Materials Research, Michigan State University, East Lansing, Michigan 48824, USA

* To whom correspondence should be addressed

TITANIUM silicalite is an effective molecular-sieve catalyst for the selective oxidation of alkanes, the hydroxylation of phenol and the epoxidation of alkenes in the presence of H_2O_2 (refs 1–3). The range of organic compounds that can be oxidized is greatly limited, however, by the relatively small pore size (about 0.6 nm) of the host framework⁴. Large-pore (mesoporous) silica-based molecular sieves have been prepared recently by Kresge *et al.*^{5–7} and Kuroda *et al.*⁸; the former used a templating approach in which the formation of an inorganic mesoporous structure is assisted by self-organization of surfactants, and the latter involved topochemical rearrangement of a layered silica precursor. Here we describe the use of the templating approach to synthesize mesoporous silica-based molecular sieves partly substituted with titanium—large-pore analogues of titanium silicalite. We find that these materials show selective catalytic activity towards the oxidation of 2,6-di-*tert*-butyl phenol to the corresponding quinone and the conversion of benzene to phenol.

We have prepared hexagonal mesoporous silica (designated HMS) and its titanium-substituted derivative (Ti-HMS) by acid hydrolysis of $Si(OC_2H_5)_4$ or of $Ti(iso-OC_3H_7)_4:Si(OC_2H_5)_4$ mixtures in alcoholic solutions using dodecylamine (DDA) as a template. This preparative strategy differs from all previously

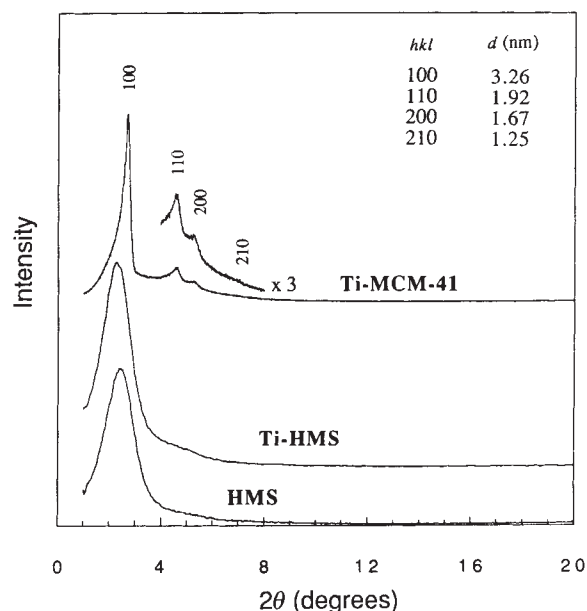


FIG. 1 X-ray powder diffraction patterns of mesoporous molecular sieves. The diffraction patterns were recorded on a Rigaku Rotaflex diffractometer equipped with a rotating anode and using $Cu K\alpha$ radiation ($\lambda = 0.15418$ nm). The *hkl* and *d* spacing given apply to Ti-MCM-41.

reported mesoporous molecular sieve syntheses insofar as a primary, rather than a quaternary, ammonium ion surfactant is used as a templating reagent. The pure silica derivative was prepared by adding a clear solution of $\text{Si}(\text{OC}_2\text{H}_5)_4$ (1.00 mol) in ethanol (6.54 mol) to a stirred solution of DDA (0.27 mol) and HCl (0.02 mol) in water (36.3 mol). Allowing the resulting gel to age for 18 h at ambient temperature afforded the crystalline templated product. An analogous procedure was used to prepare Ti-HMS from $\text{Ti}(\text{iso-OC}_3\text{H}_7)_4$ and $\text{Si}(\text{OC}_2\text{H}_5)_4$ at 1:100 molar ratio, except that a 6.5:1.0 molar mixture of $\text{C}_2\text{H}_5\text{OH}:(\text{CH}_3)_2\text{CHOH}$ was used in place of ethanol. Also, the replacement of DDA in our ambient-temperature procedure by the customary quaternary ammonium ion template $[\text{C}_{16}\text{H}_{33}\text{N}(\text{CH}_3)_3]^+$ (CTMA) (with counterion Br^-) afforded a titanium-substituted derivative of MCM-41 silica with crystallographic and mesoporous properties in accord with those reported previously for the pure silica analogue⁵⁻⁷. Before the measurements to be described below, all samples were calcined in air at 923 K for 4 h to remove the structurally incorporated template.

As shown in Fig. 1, the X-ray powder diffraction patterns for HMS and Ti-HMS are nearly identical, with both materials exhibiting a single diffraction peak corresponding to ~ 3.8 nm. An intense low-angle reflection also is a characteristic feature of Ti-MCM-41, but owing to the long-range hexagonal order weaker 110, 200 and 210 reflections are observed in the 2θ range from 4.0 to 7.0°. The uniform mesoporous structure of Ti-MCM-41 is evident in the transmission electron microscope (TEM) lattice image shown in Fig. 2a. Higher-order Bragg reflections are not resolved in the X-ray diffraction patterns of HMS and Ti-HMS (compare Fig. 1). Similar 'single reflection' MCM-41-type products, obtained by quaternary ammonium template synthesis and possessing substantial amounts of a hexagonal mesopore phase, also have been reported⁵. Electron

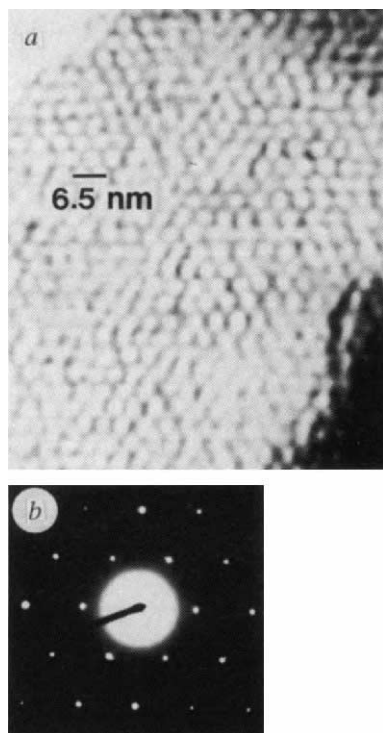


FIG. 2 a, Transmission electron micrograph (TEM) of the hexagonal pore structure of Ti-MCM-41. b, Electron diffraction pattern of Ti-HMS. The TEM image was obtained on a JEOL 100CX using an accelerating voltage of 120 kV and 20 μm objective lens aperture; the electron diffraction pattern was taken at 100 kV using 20 μm diffraction aperture.

TABLE 1 Properties of mesoporous molecular sieves and related materials

Sample	Specific surface area ($\text{m}^2 \text{g}^{-1}$)	2,6-DTBP oxidation		Benzene oxidation	
		Conversion* (%)	Quinone selectivity (%)	Conversion* (%)	Phenol selectivity (%)
Ti-HMS	1,031	83	>95	37	>95
Ti-MCM-41	1,345	20†	>98	68	>98
HMS	1,120	6.8	50	—	—
TS-1	398	6.5	>95	31	>95
TSG	7	12	67	0	0
TiO_2	20	14.5	>95	—	—

Specific surface areas and catalytic oxidation properties of the mesoporous molecular sieves (Ti-HMS, Ti-MCM-41 and HMS) and related materials (see text). Reaction conditions for these compounds are given in the text.

* Blank reactions in the absence of any catalyst gave conversions of 6.8% and 0%, respectively, for 2,6-DTBP and benzene oxidation.

† A conversion of 96% was observed after 18 h reaction time, whereas the corresponding blank reaction in the absence of catalyst gave a conversion of 16%.

micrographs of HMS and Ti-HMS reveal a very fine particle morphology (<40 nm). Thus, the diffuse scattering at $\sim 5^\circ$ for these two materials may arise from broadening of $hk0$ reflections due to finite size effects. But the electron diffraction patterns (compare Fig. 2b) exhibit well-defined hexagonal maxima, clearly establishing a crystallographic symmetry analogous to MCM-41 phases.

The regular mesoporous structures of HMS, Ti-HMS and Ti-MCM-41 are manifested in their sorption properties. Figure 3 presents the N_2 adsorption-desorption isotherms and the corresponding Horvath-Kawazoe⁹ pore size distribution curves for all three materials. Included for comparison are the corresponding plots for microporous titanium silicalite, TS-1, as prepared by the method of Tangaraj *et al.*¹⁰ and nonporous TiO_2 . In accord with the N_2 sorption properties of MCM-41 silicas⁷, the abrupt steps in the $P/P_0=0.20-0.35$ region and the corresponding maxima in the pore distribution curves indicate the existence of uniform mesopores in the size range 2.6–3.0 nm. Although the template chain is shorter for DDA than CTMA, HMS and Ti-HMS have slightly larger pore sizes than Ti-MCM-41, suggesting that the two surfactants adopt different conformations during synthesis. The crystallographic mesopore volumes for Ti-HMS and Ti-MCM-41 are very similar (0.68 and 0.69 ml g^{-1} , respectively), but the two materials differ greatly in textural (interparticle) mesoporosity, as indicated by a comparison of the hysteresis loops beyond $P/P_0=0.8$. For Ti-HMS the textural mesopore volume in the 10–30 nm range is 1.11 ml g^{-1} compared with 0.03 ml g^{-1} for Ti-MCM-41. These differences in textural porosity reflect the crystallographic distinctions indicated in the X-ray diffraction patterns of these materials and emphasize the discriminating templating properties of primary and quaternary ammonium ion surfactants. Stucky and co-workers¹¹ have proposed that the templating of hexagonal mesoporous structures proceeds through lamellar assemblies of surfactant molecules and not by direct templating around rod-shaped micelles, as originally suggested^{6,7}. It appears that the way such lamellar assemblies rearrange into hexagonal structures depends on the structure of the polar head group, as well as on the length of the surfactant chain.

Today's environmental concerns demand the streamlining of the catalytic processes for the production of fine chemicals. Both Ti-HMS and Ti-MCM-41 offer great promise in contributing to this need by complementing the catalytic oxidation chemistry of TS-1. Two peroxide oxidation reactions, namely the conversion of 2,6-di-*tert*-butyl phenol (2,6-DTBP) to the corresponding quinone and the hydroxylation of benzene to phenol, were used to demonstrate their exceptional efficiency. The catalytic experiments were conducted at 335 K for 2 h using 100 mg of catalyst, 5 mmol of 2,6-DTBP or 10 mmol of benzene as substrate,

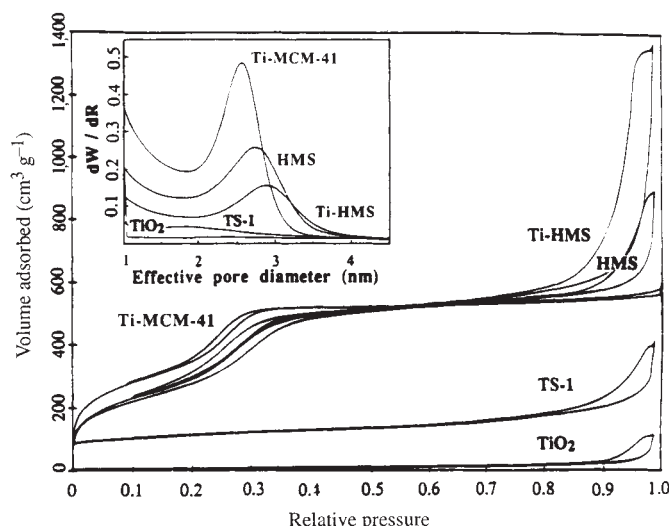


FIG. 3 Nitrogen adsorption-desorption isotherms for mesoporous molecular sieves (Ti-HMS, HMS and Ti-MCM-41), microporous titanium silicalite (TS-1), and nonporous anatase (TiO₂). (Relative pressure is P/P_0 where P is the equilibrium pressure of the adsorbate and P_0 is the saturation pressure of the adsorbate at the temperature of the adsorbent, volume adsorbed is at STP.) Insert: The corresponding Horvath-Kawazoe pore size distribution curves. (dW/dR is the derivative of the normalized adsorbate (nitrogen) volume adsorbed with respect to the pore diameter of the adsorbent.) Before measurement, each sample was heated overnight at 423 K and 10^{-6} torr. The isotherms were measured at 77 K on a Coulter Omnisorp 360CX Sorptometer using standard continuous adsorption procedures.

0.13 mol acetone and 29 mmol of 30% aqueous H₂O₂. The catalytic results, along with N₂ BET surface areas¹², are compared in Table 1 with those for HMS, TS-1, an amorphous titanium silicate gel (TSG) and bulk TiO₂ (anatase). The TSG was obtained by preparative methods equivalent to a Ti-HMS synthesis, but in the absence of template. Efforts to prepare mesoporous TiO₂ by templated synthesis gave only the bulk anatase phase.

As in the case of TS-1, both Ti-HMS and Ti-MCM-41 are effective catalysts for the direct hydroxylation of benzene. All three materials circumvent the need for a time-consuming pre-alkylation step¹³ to achieve the oxidation of the aromatic ring. But the results for 2,6-DTBP emphasize the importance of Ti-HMS and Ti-MCM-41 mesoporosity in achieving the oxidation of larger substrates. 2,6-DTBP cannot be accommodated by the micropore structure of TS-1. Owing to the absence of appreciable surface areas and/or Ti site isolation, amorphous TSG and bulk TiO₂ are equally inefficient for the oxidation of 2,6-DTBP, as well as for benzene. The general unsuitability of TSG-type materials for peroxide oxidations of aromatics also has been reported recently by other workers¹⁴. Thus, Ti-HMS and Ti-MCM-41 are important complements to TS-1 in extending peroxide oxidation chemistry to large aromatic substrates.

We note that Ti-HMS is more active than Ti-MCM-41 for 2, 6-DTBP oxidation, but the reverse is true for the oxidation of benzene. The hydroxylation of aromatic rings over both catalysts presumably occurs by a common mechanism, perhaps a peroxotitanate pathway as postulated for TS-1 (ref. 3). However, access to the framework titanium sites of Ti-HMS by the larger substrate is probably facilitated by the exceptionally high textural mesoporosity of this catalyst. In contrast, the somewhat smaller framework pore structure of Ti-MCM-41, along with the lack of appreciable textural mesoporosity for this material, probably contributes to its lower reactivity toward 2,6-DTBP. Diffusional effects in these materials will be the subject of future studies. □

Received 29 December 1993; accepted 31 January 1994.

1. Taramasso, M., Perego, G. & Notari, B. US Patent No. 4410501 (1983).
2. Huybrechts, D. R. C., De Bruycker, L. & Jacobs, P. A. *Nature* **345**, 240-242 (1990).
3. Notari, B. *Structure-Activity and Selectivity Relationships in Heterogeneous Catalysis* (eds Grasselli, R. K. & Sleight, A. W.) 243-256 (Elsevier, Amsterdam, 1991).
4. Flanigen, E. M. et al. *Nature* **271**, 512-516 (1978).
5. Kresge, C. T., Leonovicz, M. E., Roth, W. J. & Vartuli, J. C. US Patent No. 5098684 (1992).
6. Kresge, C. T., Leonovicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710-712 (1992).
7. Beck, J. S. et al. *J. Am. chem. Soc.* **114**, 10834-10843 (1992).
8. Inagaki, S., Fukushima, Y. & Kuroda, K. *J. chem. Soc. chem. Commun.* **8**, 680-682 (1993).
9. Horvath, G. & Kawazoe, K. *J. J. Chem. Engng Jap.* **16**, 470-475 (1983).
10. Tangaraj, A., Kumar, R., Mirajkar, S. P. & Ratnasamy, P. *J. Catal.* **130**, 1-8 (1991).
11. Monnier, A. et al. *Science* **261**, 1299-1303 (1993).
12. Sing, K. S. W. et al. *Pure appl. Chem.* **57**, 603-619 (1985).
13. Sheldon, R. A. *New Developments in Selective Oxidation* (eds Centi, G. & Trifiro, F.) 1-32 (Elsevier, Amsterdam, 1990).
14. Neumann, R., Chava, M. & Levin, M. *J. chem. Soc. chem. Commun.* **22**, 1685-1687 (1993).

ACKNOWLEDGEMENTS. The partial support of this research by the US NSF (research assistantship to P.T.T.), and by the NIH (research assistantship to M.C.) is acknowledged.

Pre-industrial atmospheric lead contamination detected in Swedish lake sediments

Ingemar Renberg*, Maria Wik Persson† & Ove Emteryd‡

* Department of Environmental Health,

University of Umeå, S-901 87 Umeå, Sweden

† Department of Ecological Botany, University of Umeå,

S-901 87 Umeå, Sweden

‡ Environmental Research Laboratory,

Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden

DESPITE evidence from Greenland ice cores for pre-industrial atmospheric trace-metal contamination^{1,2} it is commonly assumed that air pollution in remote areas is a recent problem caused by industrial activities, fossil-fuel burning and emissions from motor vehicles. Here we report analyses of lake sediments from Sweden showing that atmospheric lead deposition increased above background levels more than 2,600 years ago. There was a small, but marked lead deposition peak about 2,000 years ago, and a more significant increase that began 1,000 years ago and accelerated during the nineteenth and particularly the twentieth centuries, with a deposition maximum at about AD 1970. Before the nineteenth century industrialization, lead concentrations in lake sediments from southern Sweden had already reached 10-30 times previous background levels as a result of atmospheric deposition. We suggest that this pre-industrial airborne pollution was derived from extensive production and use of lead in Europe, starting with the Greek and Roman cultures^{3,4}. The cumulative deposition from anthropogenic sources in pre-industrial times (~600 BC to AD 1800) was at least as large as the cumulative deposition during the industrial period (AD 1800 to the present).

Mobility of lead in soil and sediments is usually very low⁵⁻⁷, and the transport of lead from catchment soils to lakes is also limited⁸. Most excess lead found in lakes with undisturbed forested catchments is derived from direct precipitation to the lake surfaces⁹. Therefore, we assume that concentrations in the sediments mirror atmospheric deposition. Although biases may occur, due to individual lake and catchment properties and to variations in sediment composition and accumulation rates over time, the large data set that we present is sufficient for making general conclusions about spatial and temporal lead-pollution history.

We used two sets of sediment cores; one set consisted of long cores between 1.5 and 4 m in length from 19 lakes, for studies of pre-industrial pollution trends (Fig. 1). The other set consisted

of 25–35-cm surface-sediment cores from 112 lakes, which were used for geographical surveys of pollution conditions in Sweden in about AD 1600–1800 (core bottoms) and in the 1980s (core tops) (Fig. 2).

The natural background concentrations of lead in the long cores from the 19 lakes varied between 2 and $15 \mu\text{g g}^{-1}$ (dry weight), but usually it was $<10 \mu\text{g g}^{-1}$. These low and stable concentrations were found at variable depths in the sediment cores due to different sediment accumulation rates in the lakes, but always $>1 \text{ m}$ down. Background concentrations were reached at depths much greater than previously assumed, and this may explain why background lead levels for southern Sweden have been considered—erroneously—to be $\sim 50 \mu\text{g g}^{-1}$ (dry weight)¹⁰. A series of contiguous sediment samples from Lilla Holmevatten and Lilla Öresjön (40 5-cm and 24 10-cm samples respectively) covering the period from 10000 BC (deglaciation) to about 3000 BC, showed very little temporal variation in lead concentrations (Fig. 3d). These results suggest that natural catchment disturbances, such as forest fires, did not cause increased influx of lead to the lakes. Neither do these data indicate large and variable atmospheric lead deposition from volcanic eruptions¹¹.

From this data set we cannot determine exactly when the lead concentration increased over the natural background level, but

the two radiocarbon dates made of the increase at Stora Skärsjön and Rudegyl (Fig. 1) gave very similar results, suggesting that it occurred $\sim 2,600 \text{ yr}$ ago. The subsequent peak that reached about five times over background concentrations, and which is detectable all over Sweden, was dated in six lakes to $\sim 2,000$ radiocarbon years before present (BP) (Fig. 1). Also, the major and steep concentration increase that occurred $\sim 1,000$ radiocarbon yr BP is synchronous (dates from 10 lakes, Fig. 1). Although there are indications of slightly lower values a few hundred years later in several southwest Swedish lakes, lead concentrations have never returned to near-background levels over the past 1,000 yr in southern Sweden.

There are several factors that suggest that the increased pre-industrial lead concentration resulted from atmospheric deposition. First, the temporal concentration changes are synchronous over a very large geographical area ($>150,000 \text{ km}^2$). Second, the spatial concentration variations over Sweden resembles present-day deposition maps of long-range transported airborne pollutants, with high values in the southern part of the country close to continental Europe, and low values in the north (Fig. 2). Third, there is no obvious relationship between the changes in lead concentration and land-use history (Table 1).

The temporal changes in lead deposition from the atmosphere recorded in these lakes agree well with the history of world lead

FIG. 1 Lead concentrations (dry weight) in pre-industrial sediments from Swedish lakes plotted against depth in the sediment cores. Note that the diagrams do not cover the industrial period (past 200–300 yr) except in four lakes (Lilla Holmevatten Härsvatten, Lilla Öresjön, Rudegyl). Sediment cores for analyses of pre-industrial lead concentrations were sampled using a Russian corer (length 1 m, diameter 8 cm) from ice in winter at the deepest point of each lake. The sediment–water interface and the first 5–10 cm of the sediment cannot be retrieved with this corer. In addition, depending on estimated sediment accumulation rates, lead concentrations in the uppermost 20–50 cm of the cores are not plotted, to ensure that the sequences presented here start below the industrial period. Cores of recent sediments covering the industrial period were taken with a freeze or a HON-Kajak corer^{15,16}. The lakes are headwater lakes, 10–30 ha in surface area, $>10 \text{ m}$ deep and oligotrophic; there are no industries and usually no fields or houses in the catchments. Organic content of the sediments as measured by loss on ignition is typically ~ 30 – 50% and does not vary much within the cores. Dried sediment samples were digested with $\text{HNO}_3 + \text{HClO}_4$ (10:1) for 2–5 h at $<130^\circ \text{C}$, and lead analyses were made using inductively coupled plasma mass spectrometry (ICP-MS). Radiocarbon dates (yr BP) are uncalibrated AMS dates of terrestrial plant remains such as seeds, leaves and needles sieved from the cores (Ua-3451–Ua-3467). Only in Härsvatten, Lilla Holmevatten and Stora Längen have bulk sediment samples been dated with AMS (Ua-2704–Ua-2707, Ua-2711, Ua-2712). Bulk samples often become slightly too old due to reservoir effects¹⁷ and this may explain why, in these lakes, the age of the major lead increase is a few hundred years older than in the other lakes. The deviation between radiocarbon dates and calendar years is small for the time period under consideration. However, a more detailed study of the lead deposition history would require correction of the radiocarbon dates and also closer sampling intervals for the lead analyses.

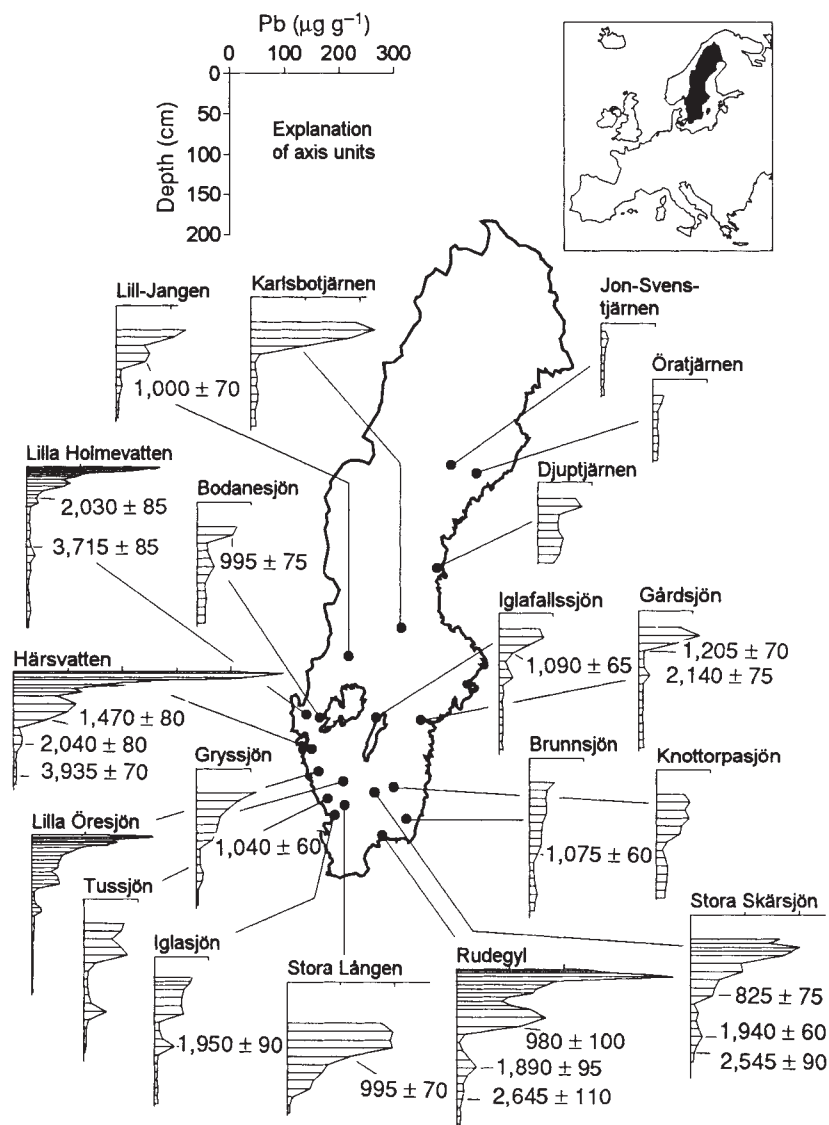


TABLE 1 Sediment cores from Swedish lakes

Lake	Difference (cm)
Härsvatten	0
Stora Skärsjön	+30
Stora Lången	+10
Lilla Holmevatten	-10
Rudegyl	0
Iglasjön	-20
Brunnsjön	+10
Gryssjön	-30
Knottorpsjön	+10
Tussjön	No pH change
Bodanesjön	No pH change

Comparison between lead pollution and land-use histories. Land-use history has previously been studied using pollen, charcoal and loss-on-ignition analyses of sediments from six of the above lakes and the pH history in 11 of the lakes with diatom analyses¹⁴. In that study, we found a correlation between land use and increased past-lakewater pH inferred from the subfossil diatom record. Here we use the pH increase as an indicator of land-use activity, and compare the depths in the sediment cores of this land-use-induced pH increase and the general lead increase dated to ~1,000 yr ago. The same cores were used for both studies. + indicates that the expansion of land use occurred above (after) the lead concentration increase, and - below (before). In most of the lakes the land-use expansion occurs either before or after the lead concentration increase, indicating that land-use is not the main cause of increased lead concentrations in the sediments.

production³ (Fig. 3). During the Roman period, the production of lead was already considerable in Europe and Asia Minor (~80,000 ton yr⁻¹, ref. 12), and it has been suggested that 5% of the mined lead was emitted into the atmosphere¹². We believe our data provide evidence for early long-range transport of airborne pollutants from continental Europe to Sweden; these lake-sediment data may also support the hypothesis of a general increase of the lead concentration in the atmosphere of the Northern Hemisphere by a factor of 5 as a result of anthropogenic activities, as early as 2,000 yr ago¹². The steep increase of lead concentration in the lake sediments about 1,000 yr ago may reflect increased production and refining of lead in northern Europe, for example in Germany^{3,13} (Fig. 3a).

Both the long cores (Fig. 1) and the map based on 112 short surface-sediment cores (Fig. 2a) show that lead pollution in southern Sweden was considerable before the industrial period. Even before the period of marked industrialization, rural pollution conditions were at a comparable level to that in large areas of northern Sweden in the 1980s (Fig. 2b), despite emissions from motor traffic and a large lead smelter on the Baltic coast of northern Sweden.

We have also used the two sediment-core sets to compare the cumulative anthropogenic lead accumulation in industrial and pre-industrial times (Fig. 2c, d). These data indicate that in the southern part of Sweden, which has been most exposed to long-range transported lead pollution, the accumulated lead deposition in pre-industrial times was as large as, or even larger than, the deposition during the comparatively short industrial time

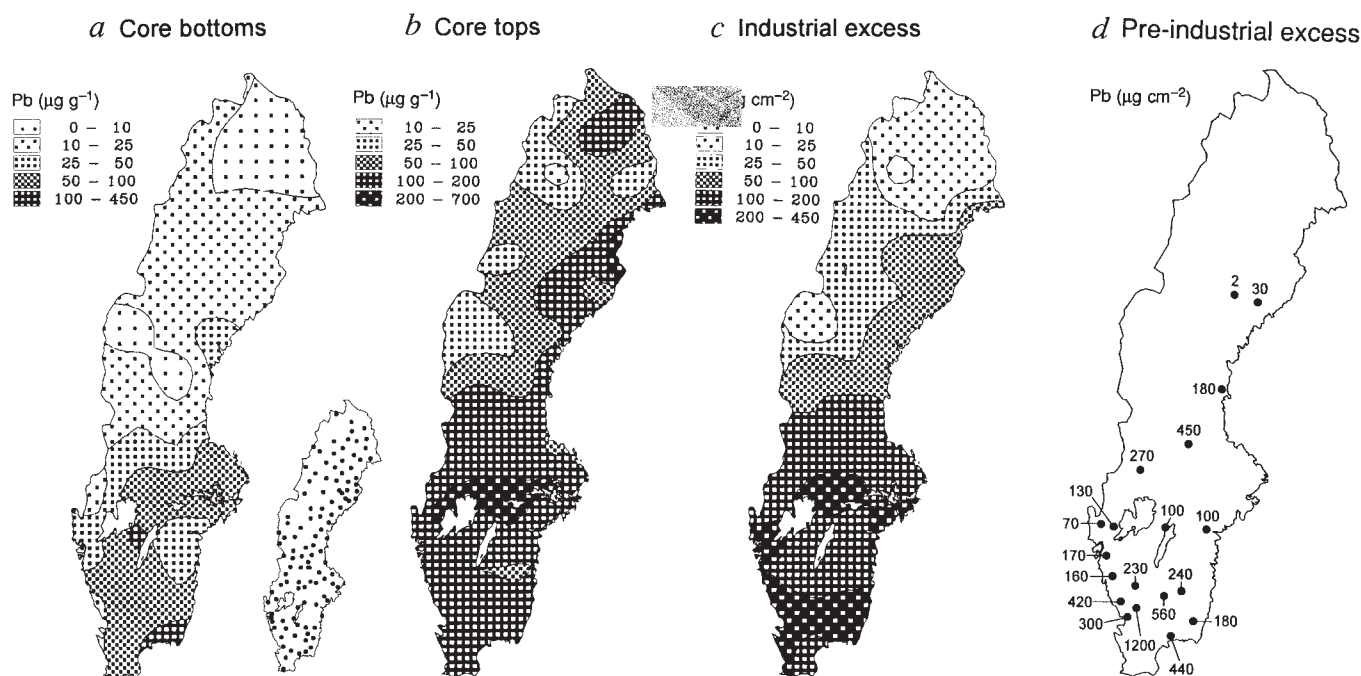


FIG. 2 a, Lead pollution conditions in Sweden about AD 1600–1800 as reflected by lead concentrations (dry weight) in core bottoms of 25–35-cm surface-sediment cores from 112 lakes. Cores were sampled with a HON-Kajak gravity corer¹⁶ in 1986. Insert map shows location of lakes. Lead analyses were made with atomic absorption spectroscopy (AAS) after HNO₃ + HClO₄ extraction. b, Lead pollution conditions in the 1980s according to analyses (Pb, µg g⁻¹, dry weight) of core tops (0–1 cm) of the sediment cores (from a). c, Excess anthropogenic lead accumulation during the industrial period in these sediment cores, estimated in the following way. At the same time that the cores were taken for analyses of core bottoms and core tops, one 25–35 cm core was also obtained from each lake. This core was handled as one sample; it was thoroughly mixed, its dry mass was determined, its lead concentration was analysed with AAS, the natural lead concentration

(10 µg g⁻¹, dry weight) was subtracted from the mean lead concentration of the total sample, the excess lead in the total sample was calculated, and this value was divided by the surface area of the coring tube to obtain the excess cumulative lead accumulation per cm⁻² on the lake bottom at the coring site. (Maps a–c were made using the Uniras computer program.) d, Excess anthropogenic lead accumulation during the pre-industrial period (about 600 BC to about AD 1800) in the long cores from the 19 lakes. These values were calculated in an analogous manner as above, although using the series of lead analyses from the long cores. These values should be regarded as guidelines; a more detailed comparison between pre-industrial and industrial excess values would require more rigorous dating, exact determination of background values for individual lakes, and analyses of cores covering both the industrial and pre-industrial periods from a larger number of lakes.

period. Although high concentrations of metals during a short period may be more deleterious to the ecosystem than low concentrations over a long period, the pre-industrial load must not be neglected because these metals are stored in the soil and can potentially be mobilized.

These results demonstrate large-scale pre-industrial lead contamination. It is very likely that atmospheric transport of

several other metals (and sulphur) also occurred. We believe our investigation challenges the commonly held view of a 'clean' pre-industrial environment. □

Received 24 November 1993; accepted 1 February 1994.

1. Murozumi, M., Chow, T. Y. & Patterson, C. C. *Geochim. cosmochim. Acta* **33**, 1247–1294 (1969).
2. Boutron, C. F. & Görlach, U. in *Metal Speciation in the Environment* (eds Broekaert, J. A. C., Güter, S. & Adams, F.) 137–151 (NATO ASI Ser. G 23, Springer, Berlin, 1990).
3. Settle, D. M. & Patterson, C. C. *Science* **207**, 1167–1176 (1980).
4. Nriagu, J. O. *Lead and Lead Poisoning in Antiquity* (Wiley, New York, 1983).
5. Dörr, H. & Münnich, K. O. *Wat. Air Soil Pollut.* **57–58**, 809–818 (1991).
6. Lazerte, B., Evans, D. & Grands, P. *Sci. tot. Envir.* **87/88**, 209–221 (1989).
7. Livett, E. *Adv. ecol. Res.* **18**, 65–177 (1988).
8. Lindberg, S. E. & Turner, R. R. *Wat. Air Soil Pollut.* **39**, 123–156 (1988).
9. Borg, H. & Johansson, K. *Wat. Air Soil Pollut.* **47**, 427–440 (1989).
10. *Bedömningsgrunder för Sjöar och Vattendrag: Klassificering av Vattenkemi samt Metaller i Sediment och Organismer* (Swedish Environmental Protection Agency, Solna, 1990).
11. Nriagu, J. O. *Environment* **32**, 7–11 (1990).
12. Shirahata, H., Elias, R. W., Patterson, C. C. & Koide, M. *Geochim. cosmochim. Acta* **44**, 149–162 (1980).
13. Forshell, H. *Theses and Papers in Archaeology* Vol B2 1–189 (Archaeological Research Laboratory, Stockholm University, 1992).
14. Renberg, I., Korsman, T. & Birks, H. J. B. *Nature* **362**, 824–827 (1993).
15. Renberg, I. *Boreas* **10**, 255–258 (1981).
16. Renberg, I. *J. Paleolimnol.* **6**, 167–170 (1991).
17. Olsson, I. U. in *Handbook of Holocene Palaeoecology and Palaeohydrology* (ed. Berglund, B. E.) 273–312 (Wiley, Chichester, 1986).
18. Stuiver, M. & Reimer, P. J. *Radiocarbon* **35**, 215–230 (1993).
19. Wik, M. & Renberg, I. *Ambio* **20**, 289–292 (1991).

ACKNOWLEDGEMENTS. We thank The Svedberg Laboratory, Uppsala University for radiocarbon dates, P. Lidstöm and E. Renberg for preparation of computer graphics, D. Lindberg, U. Segerström and J.-E. Wallin for fieldwork assistance, C. C. Patterson for permission to use Fig. 3a and N. J. Anderson for comments on the manuscript. This work was supported by the Swedish Environmental Protection Agency and the Swedish Natural Science Research Council.

Sulphur output and magma degassing budget of Stromboli volcano

P. Allard*, J. Carbonnelle††, N. Métrich§, H. Loyer†† & P. Zettwoog†||

* Centre des Faibles Radioactivités, CNRS-CEA, 91198 Gif sur Yvette, France

† Commissariat à l'Energie Atomique, IPSN, 91191 Saclay, France

§ Laboratoire Pierre Süe, CNRS-CEA, 91191 Saclay, France

STROMBOLI volcano in the Aeolian islands has been erupting continuously for more than 2,000 years¹, and probably as many as 5,000, following a major flank collapse^{2,3}. Here we describe airborne measurements of the plume flux of SO₂ during 1980–93, which show that the volcano emits very large amounts of gas, mostly by open-conduit degassing between explosive outbursts, while exuding little basalt. Microprobe analysis of sulphur in the K-rich (shoshonitic) basalt, along with data for primitive basalts in the region^{4,5}, suggests that the time-averaged SO₂ flux is produced by intrusive degassing of 0.01–0.02 km³ yr⁻¹ of magma, 100–200 times more than the volume erupted. Over 5,000 years, this rate implies that 50–100 km³ of intruded basalt would have been degassed, suggesting either that the volcanic pile has grown substantially by intrusion⁶ or, more probably, that a large magma storage system is emplaced at a shallow level within the crustal basement. Our results indicate that Etna and Stromboli alone provide 10% of the global budget of volcanic SO₂.

Stromboli is a strato-volcano ~3 km high, built upon the Tyrrhenian sea floor and rising 924 m above sea level⁷. Its typical

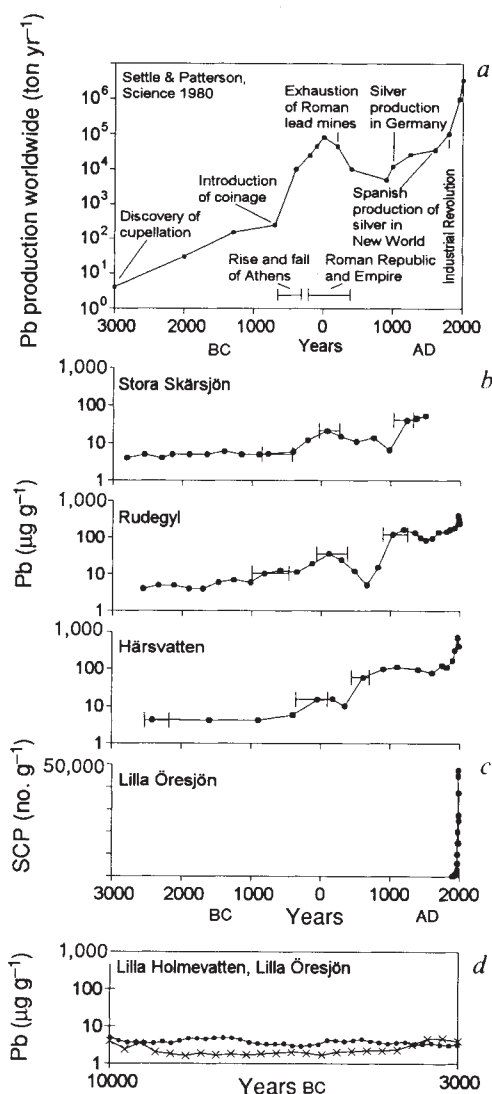


FIG. 3 a, Diagram from Settle & Patterson (ref. 3, © AAAS) showing the lead production in the world over the past 5,000 yr. These production figures are estimated from various historical sources. b, Lead concentrations (dry weight) in the sediment cores from the lakes (shown in Fig. 1) that have three radiocarbon dates, each plotted against time (calibrated radiocarbon dates¹⁸, error bars 2 σ). Both a and b are plotted on a logarithmic scale that exaggerates pre-industrial production and concentration values, but facilitates comparison of the two diagrams. c, To illustrate the temporal difference between the pre-industrial and the industrial period, the concentration (number per gram, dry weight) of spheroidal carbonaceous particles (SCP) in a ²¹⁰Pb-dated sediment core from Lilla Öresjön have been plotted. These fly-ash particles are characteristic of oil and coal combustion and are good indicators of deposition of airborne pollutants derived from fossil-fuel burning¹⁹. d, Lead concentrations (dry weight) in contiguous samples covering the period ~10000 bc to 3000 bc from Lilla Holmevatten and Lilla Öresjön.

†|| Present addresses: † MSIS, ZAC de Courcelle, 91196 Gif/Yvette, France; || Certac, 78610 Auffargis, France.

TABLE 1 SO₂ flux from Stromboli

Date	6 October 1980	21 September 1984	31 May 1991	1 June 1991	3 June 1991	6 June 1993
Local time	13:59–14:45	12:30–13:20	15:30–16:10	12:10–13:00	11:05–12:22	13:44–14:40
Wind speed (m s ⁻¹)	25	6.0	8.2	12.0	11.3	5.7
Volcanic activity level*	High	Medium	Low	Low	Medium	Low
No. of traverses	8	10	6	12	11	14
Flux values† (t d ⁻¹)	1,370	690	460	430	715	260
	1,180	1,130 (E)	575	1,315 (E)	1,555 (E)	250
	1,450 (E)	890	670	535	670	220
	1,540 (")	820	500	325	1,640 (E)	380
	1,730 (E)	710	300	300	960	350
	1,050	920	690	425	1,795 (E)	410 (E)
	1,670 (E)	730		320	1,710 (")	295
	1,360	1,010 (E)		450	1,690 (")	315
		1,230 (E)		390	930	370
		845		300	775	315
				340	805	330
				365		550 (E)
						370
						340
Mean‡	1,420 ± 215	900 ± 170	530 ± 145	460 ± 105	1,205 ± 465	340 ± 80
Non-eruptive mean	1,240 ± 135	800 ± 85	530 ± 145	380 ± 70	810 ± 105	320 ± 50
Eruptive mean	1,600 ± 110	1,120 ± 90	—	1,315	1,680 ± 80	480 ± 70

Fluxes were measured by operating COSPEC onboard of a twin-engine laboratory aircraft (BN IIA), during traverses at 300–500 m height below the volcanic plume and 0.5–5 km distance from the summit crater. Wind (plume) speed is from either proximal airport data (1980) or the aircraft navigation system (1984–93). The uncertainty on mean fluxes ranges from 15% (1993) to 25% (1980). Other available data do not distinguish between eruptive and non-eruptive emissions. Stoiber *et al.*¹⁴ reported a mean SO₂ flux of 135 t d⁻¹ in 1975, from ground-based COSPEC measurements. Using COSPEC from a helicopter, Caltabiano and Romano measured one value of 400 t d⁻¹ in November 1987 (ref. 19) and 360 t d⁻¹ in April 1993 (personal communication), which is similar to our 'low mean' for the non-eruptive flux. Variations of the volcanic gas output, but also differences in the experimental procedures, may lead to differing fluxes. Owing to both spectrometric constraints and Stromboli's steep topography, COSPEC traverses with an aircraft provide the most reliable conditions for measuring the gas output of this volcano. On a time-averaged basis, the non-eruptive emission supplies most of the SO₂ output, the contribution of explosive outbursts being episodic and of short duration (see text).

* According to the frequency and size of explosions over the days of measurement or more.

† (E) indicates eruptive plume from single explosion or (") same explosion.

‡ Mean of all observations, ±1σ.

activity consists of mildly explosive jets of incandescent gas and lava clots from summit crater vents (on average, 5–6 events of 5–10 s each per hour; refs 8–11 and authors' own observations), interspersed by quiescent degassing with puffs and flames. Short-lived episodes of more violent explosions and lava flows also happen periodically¹², the last in 1985 (ref. 13).

Both quiescent and explosive degassing sustain a volcanic plume, the SO₂ yield of which has been measured only once¹⁴ (in 1975) before our study. In 1980–93, we measured the SO₂ plume flux on several occasions, between and during explosions (Table 1), using remote correlation spectrometry (COSPEC) from an aircraft^{15,16}. The mean flux during quiescent intervals was found to vary from 320 to 1,200 tons per day (t d⁻¹), depending on the intensity of volcanic activity, and to be ~800 t d⁻¹ during medium activity. The peak values corresponding to eruptive gas clouds indicate a 40–400% relative flux increase during explosions, with a unit rate of (5–19) × 10⁻³ t s⁻¹ of SO₂ that is coherent with an estimate of (24–30) × 10⁻³ t s⁻¹ for total gas⁹. But due to their episodic nature and short duration (≤24 min of eruption per day as a whole), these events provide only a few per cent of the time-averaged SO₂ output; the non-eruptive contribution is predominant. This is likely to be the case also for the other main volatile species. Thus the volatile output of Stromboli is best approximated by the quiescent, not explosive, emissions, a fact that has been overlooked previously^{9,17}. From ground-based COSPEC measurements in 1975, Stoiber *et al.*^{14,18} estimated 4.9 × 10⁴ t yr⁻¹ of SO₂ from the volcano. Our results and those of others¹⁹ for low to medium activity in 1980–93 suggest a mean output within the range (1.5–2.9) × 10⁵ t yr⁻¹. Stromboli thus produces 5–10 times less SO₂ than nearby Etna^{19,20}, but both Italian volcanoes alone supply as much as 10% of the worldwide volcanic yield of this gas^{18,21}. Inferring the mean fluxes of H₂O, CO₂, HCl and HF from the

composition of Stromboli's quiescent plume and fumarolic emissions (Table 2), we assess its total gas discharge at (6–12) × 10³ t d⁻¹ (±50%). The output of solid material during 'normal' explosions is variable but low^{9,22}—between 15 and 150 t d⁻¹ at mean frequency—and the rate of lava flow averaged over the past 100 yr^{12,13} is only 445 t d⁻¹ (~165 m³ d⁻¹) (the episodic supply of more violent explosions is poorly constrained). The total gas-to-solid mass ratio for time-averaged emissions is thus 10–26 (range: 5–40), compared to 2–16 during explosions⁹. Hence we emphasize that Stromboli emits much more gas than lava, and that volatiles contribute more importantly than expected²³ to its thermal budget (Table 2).

The SO₂ clearly rises through molten lava in the vents, even though its ultimate origin is uncertain. Isotopic δ³⁴S values of +3.5‰ for sulphate adsorbed on ashes from the 1985 eruption¹³ support a mantle derivation of the precursory gas and, together with other geochemical data³, leave little room for the possibility of magmatic assimilation of either sea water (~+20‰) or evaporitic sulphur from the crustal basement²⁴. In order to determine the magma sulphur content, we analysed melt inclusions in olivine and clinopyroxene phenocrysts of recent pyroclasts (Table 3). The pyroxenes show complex oscillatory zoning of diopside and augite (refs 12, 25), crystallized at 1,145 ± 5 °C and 1,110 ± 8 °C, respectively. The most primary and/or less degassed melt inclusions, in diopside and olivine, contain 0.085–0.10 wt% S, as also found in primitive oceanic basalts²⁶. However, correcting for their evolution with respect to the bulk lava suggests only ~0.06 wt% S in the parent liquid (Table 3). This is a minimum, taking account of potential sulphur loss by sulphide immiscibility and degassing before melt entrapment. As an upper limit, we consider here that Stromboli parent magma may contain as much sulphur as primitive basalts from nearby Etna⁴ and Vulcano⁵, that is ~0.3 wt%.

TABLE 2 Total gas output from Stromboli

Species (X)	H ₂ O	CO ₂	SO ₂	HCl	HF	Total
Weight ratio X/SO ₂						
(1) Fumarole, 201 °C	7.5	5.0	1.0	0.21	—	
(2) Fumarole, 390 °C	5.2	4.6	1.0	—	—	
(3) Fumarole, 410 °C	11.1	13.4	1.0	0.10	0.008	
(4) Dilute plume	—	6.2	1.0	0.17	0.002	
Mean ratio	7.9	7.3	1.0	0.16	0.005	
Mean flux (10 ³ t d ⁻¹)	3.2–6.3	2.9–5.8	0.4–0.8	0.06–0.13	0.002–0.004	6.1–12.3
Standard deviation	1.2–2.5	1.6–3.2	0.1–0.2	0.02–0.04	0.001–0.003	2.9–5.9

Total gas output was inferred from the time-averaged SO₂ flux, and the weight ratios of H₂O, CO₂, HCl and HF to SO₂ in available analyses of both fumarolic emissions and the quiescent volcanic plume (direct sampling of the magmatic gas phase is prevented by hazardous access to the vents). Locations of samples as follows. (1) Southwest crater, June 1964. Data from ref. 38, combined with H₂O/CO₂ ratio from ref. 8. (2) Northeast crater rim, June 1990 (evacuated bottle, this study). (3) Northeast crater rim, May 1991 (soda bottle)³⁹. (4) Mean of five best samples taken from above the crater platform (Pizzo Sopra La Fossa, June 1990; this study). Non-eruptive plume pumped into PTFE bags and analysed with Dräger tubes (± 5 –15%). CO₂ concentrations corrected from the air background. H₂S undetected. H₂O and CO₂ in crater-rim fumaroles were observed to increase during explosions⁸ (due to higher gas flow and lower air dilution), while preserving a steady ratio. The Cl/S ratio in both the plume and fumaroles is quite similar to Cl/S in the volatile fraction released from the melt (Table 3). Nevertheless, because of any possible SO₂ removal, the lower ratios and, thus, the lower fluxes may be more representative. Using the lower bound of 3×10^3 and 1.5×10^3 t d⁻¹ for the output of total gas (F) and water (F_w), respectively, we evaluate at 6 MW the minimal heat flux Q carried by the gas phase: $Q = F_c \Delta T + F_w L$. The gas mixture, assumed to be ideal, is allowed to cool from 1,000 °C to ambient temperature, with a specific heat capacity (c) of 1.88 J g^{-1} per degree C, and a latent heat of vapour condensation (L) of $2.26 \times 10^3 \text{ J g}^{-1}$. This figure is one order of magnitude greater than previously inferred from only explosions²³.

Residual sulphur in the matrix glass ($\leq 0.01 \text{ wt}\%$) implies a loss of ≥ 0.05 to possibly $\geq 0.29 \text{ wt}\%$ S during degassing. Scaling the mean SO₂ output to the highest latter quantity, we calculate a minimum magma degassing rate of $(2.5\text{--}5.0) \times 10^7 \text{ t yr}^{-1}$, that is $\sim 0.01\text{--}0.02 \text{ km}^3 \text{ yr}^{-1}$ for a basalt density of 2.7 t m^{-3} : this is 100–200 times greater than the eruption rate of both tephra and lava. In other words, to supply the SO₂ output the erupted magma should originally contain 30–60 wt% S, which is unrealistic. Thus, the gas derives almost entirely from unobserved magma.

Is this magma extruded under sea water? This is conceivable, because the volcano is essentially submarine: its aerial cone has a volume of only 4 km^3 , compared to $\sim 230 \text{ km}^3$ for the whole edifice. However, there is no evidence of such an outflow which, apart from being voluminous, ought to be also regular to permit stable lava ponding at the summit. So, we conclude that most emitted SO₂ (and other volatiles) derive from the open-conduit degassing of non-erupted (intruded) magma. This means that SO₂-bearing gas bubbles must rise separately from the melt

above some depth, before bursting either quietly (bubbly flow) or explosively (slug flow) at the surface, as previously modelled^{9,10,27}. Note that bubbly flow, although more tenuous, is quantitatively the main process of gas release. ‘Excessive’ SO₂ emission, due to heterogeneous degassing of differentiated and/or mafic intrusive magma bodies²⁸, is a common feature at andesitic–dacitic arc volcanoes^{28–32}. Here we verify the same phenomenon at a persistently active basaltic volcano which, in that respect, behaves almost like a lava lake. Masaya (Nicaragua)³³ and, to a lesser degree, Etna³⁴ are other known basaltic analogues.

The steady eruptive regime of Stromboli for thousands of years requires a continuous supply of magmatic heat and gas to its upper conduits. Taking the modern degassing rate as representative of the past rate provides new insight into the (poorly elucidated) magma feeding system of this remarkable volcano (Fig. 1). At the minimum rate of $\sim 0.01\text{--}0.02 \text{ km}^3 \text{ yr}^{-1}$ estimated here, 20–40 km³ of magma could have been degassed since the Antiquity, and 50–100 km³ since the onset of the present eruptive

FIG. 1 Conceptual model of Stromboli's magma feeding system. The volcano is 3 km high, mostly submarine, and built upon continental crust $\sim 18 \text{ km}$ thick²⁴. Its persistent eruptive activity may be favoured, locally, by gravitational instability of its northwestern flank (Sciara del Fuoco), allowing the main southwest–northeast volcano-tectonic fracture zone⁷ of the edifice (along which the active vents are aligned) to remain open and filled with magma (dyke-sheet intrusion?). Its upper eruptive conduits may be narrow (perhaps 10–20 m wide in total¹⁷) and not deeper than 200–400 m (refs 9, 17), thus being able to contain at most $\sim 10^5 \text{ m}^3$ of magma. Removal of such an amount by explosions and lava flows requires $\sim 1 \text{ yr}$, on average, during which $\geq 10^7 \text{ m}^3$ of magma are degassed for sulphur (see text). At the same rate, 50–100 km³ of magma could have been degassed over the past 5,000 yr, only a small proportion of which was erupted. So, in addition to possible intrusive growth of the edifice⁵, whose total volume is $\sim 230 \text{ km}^3$, a large reservoir of intruded basalt must extend within the local crust. The weak and shallow seismicity^{19,20,35} and thermal balance calculations¹⁷, together with the solubility behaviour of sulphur (whose exsolution may initiate within 1 km under the base of the volcano; see text), suggest that this reservoir, filled with convecting shoshonitic basalt (SHO), is emplaced and possibly growing at shallow crustal level. It sustains a heterogeneous flow of SO₂ through within-pile conduits, filled with liquids progressively more evolved and more degassed upward, represented by analysed melt inclusions in pyroxene crystals (Table 3). It may be replenished with more primary (HKCA, high-K calc-alkaline³) magma and gas (CO₂) from a deeper chamber, lying at $\sim 10\text{--}14 \text{ km}$ depth, whose existence was postulated on basis of petrological data³ and seismic velocity anomalies¹². Vertical and horizontal scales representative.

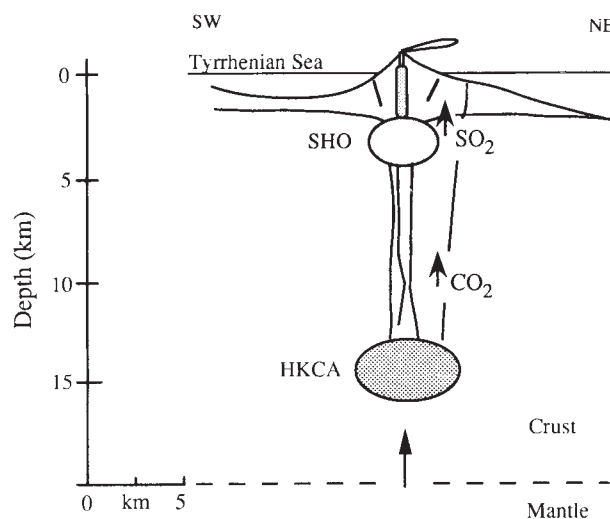


TABLE 3 Abundances of S and Cl in crystal melt inclusions and groundmass of Stromboli's pyroclasts

	June 1990			June 1980			Dec. 1985 eruption*	
	Melt inclus. in olivine†	Melt inclus. in augite	Matrix glass	Melt inclus. in diopside‡	Melt inclus. in augite§	Matrix glass	Bulk lava flow	Melt inclus. in augite
SiO ₂	51.54	52.69	51.41	53.02	52.46	52.84	50.22	52.26
TiO ₂	1.30	1.54	1.42	1.23	1.21	1.47	1.02	1.61
Al ₂ O ₃	16.84	15.36	15.48	15.30	15.61	15.69	18.03	15.66
FeO	10.02	9.53	9.69	9.08	8.56	9.37	7.22	9.40
MnO	0.18	0.17	0.19	0.18	0.16	0.19	0.17	0.21
MgO	3.73	3.33	3.80	4.37	3.74	3.55	5.63	3.38
CaO	7.26	7.00	7.68	8.67	7.21	7.12	11.27	6.64
Na ₂ O	3.62	3.26	3.40	2.88	3.58	3.60	2.69	3.08
K ₂ O	3.39	3.76	3.91	2.17	4.16	4.20	2.21	5.27
P ₂ O ₅	0.83	0.95	1.03	0.50	1.00	0.97	0.43	—
Sum (%)	98.72	97.60	98.00	97.40	97.69	99.00	99.36	97.54
S (p.p.m.)	1,000 (40)	<100	<100	880 (100)	<100	<100	—	—
Cl (p.p.m.)	2,030 (100)	1,170 (40)	1,170 (40)	1,525 (80)	1,170 (60)	915 (50)	—	—
No. of analyses	3	4	8	5	4	7	—	—
K ₂ O/Na ₂ O	0.94	1.15	1.15	0.75	1.2	1.2	0.8	1.7
K ₂ O/P ₂ O ₅	4.1	4.0	3.8	4.5	4.1	4.5	5.1	—

Analyses with a Camebax electron microprobe (Service Camparis, Paris VI University), in conditions described in ref. 4 (with 15 kV, 30 nA and 60s count rate for S and Cl in 1990 samples). 1 σ standard deviation on S and Cl contents (in p.p.m.) is given in brackets. Note that melt inclusions in augite are depleted in volatiles and compositionally equivalent to the matrix glass (for example, K₂O/Na₂O > 1), thus representing late entrapment of an evolved and degassed basaltic liquid which experiences the upper plumbing system of the volcano. Their constant composition through time (this work and refs 3, 10, 20) indicates steady-state crystallization of a homogeneous magma. Melt inclusions in diopside are more primary (richer in Ca and Mg) and are richer in both S and Cl, but are also evolved with respect to the bulk shoshonitic lava—as a consequence of 30–35% previous crystal fractionation. Correcting their sulphur content for this effect leads to only 0.06 wt% S in the original liquid. This is a minimum, however, as their sulphur content fits with the FeO–S saturation trend for S²⁻ in basaltic glasses²⁶, allowing for possible sulphur loss by degassing and/or sulphide immiscibility during pre-entrapment differentiation (although we observed sulphide globules only in augite). Moreover, according to data in ref. 40, 40% of sulphur in the parental shoshonitic liquid (log pO₂ $\approx$ -7 at 1,200 °C)¹² might have occurred as sulphate. The sulphur content of primitive basalts from nearby Etna⁷ and Vulcano⁵ suggests a possible upper concentration of $\sim$ 0.3 wt% in the parent magma of Stromboli. Combining this value with both the residual content of the matrix glass and the mean SO₂ output allows us to assess the minimal rate of magma degassing (see text).

* De Fino et al.¹³. Data for Sr 6 lava flow (recalculated with all Fe as FeO) and a crystal melt inclusion in Sr 3 lava.

† Fo 71.0–75.0%.

‡ Wo 44.1–45.4%, Fs 7.6–5.1%, En 48.3–49.5%.

§ Wo 40.2–42%, Fs 17.2–12%, En 42.6–46.0%.

|| Detection limit.

cycle, most of which has degassed intrusively. As the aerial cone has preserved a steady morphology⁷, we are faced with two possibilities. (1) The submarine edifice has experienced a recent phase of rapid endogenous growth⁶. But the magma volumes involved would imply it increased in size by 25–75% in the past 5,000 yr, which is hard to believe. (2) A large magma storage system, several tens of cubic kilometres at least, underlies the volcano. This idea is consistent with a steady thermal supply¹⁷, the uniform chemistry of historical basalts^{3,12,13} and the location of the island on a residual gravity high¹². Direct feeding of the present-day activity of Stromboli by a deep magma chamber, sited at $\sim$ 10–14 km depth within the local crust¹², was previously suggested on basis of petrological data³ and seismic velocity anomalies²⁴. However, this is difficult to reconcile with other evidence, such as the weak and shallow seismicity^{12,13,35} and thermal balance calculations¹⁷, which rather point to a shallow magma body. Moreover, this would require that SO₂ exsolution initiates under much greater confining pressure than generally observed. Although it is a complex process²⁶, possibly facilitated²⁸ by earlier bubbling of less soluble CO₂, sulphur exsolution from anhydrous basaltic melts usually occurs at $\leq$ 10 MPa for S contents $\leq$ 0.1 wt% (refs 36, 37) and was found to become vigorous at around 100 MPa at Etna⁴, whose basalts are particularly rich in S (0.3 wt%), CO₂ and water^{4,34}. The latter value corresponds to a magma column height of 3.5 km. Therefore, unless Stromboli magma (density 2.7 t m⁻³) contains still more sulphur than Etna basalts, SO₂ may start to bubble within $\sim$ 1 km under the base of the volcano. In order to sustain the SO₂ flux, undegassed magma must be continuously renewed to such a depth. Now, as the amounts of degassed magma cannot be simply accommodated by intrusive growth of the edifice, and

thus should be somehow removed downwards, we infer that a reservoir of convecting shoshonitic basalt, connected to the eruptive vents, is actually emplaced at upper crustal level (Fig. 1). This reservoir may be replenished with more primary magma and gas (CO₂) from a deeper chamber and may have been growing for 5,000 yr. Its degassing sustains a differential flow of SO₂ through the inner plumbing system of the edifice which, according to glass inclusions in augite, is mainly filled with evolved (cooler) and degassed basaltic liquid. This more viscous liquid may inhibit the eruption of lava⁶ and favours the upward accumulation of rising gas bubbles in a foam layer²⁷, which recurrently disrupts as overpressure is reached (explosions). As the oscillatory zoning of pyroxene (and plagioclase) phenocrysts indicates, it repeatedly mixes at deeper levels with hotter and less degassed liquid (diopside melt inclusions). Chemical and isotopic analysis of sulphur in more primitive products of Stromboli are required to improve this model, which could be tested by seismic tomography. □

Received 8 November 1993; accepted 11 February 1994.

- Mercalli, G. *Atti. Soc. Ital. Sci. nat.* **24**, 105–135 (1881).
- Gillot, P. Y. thesis, Univ. Paris sud. (1984).
- Francalanci, L., Manetti, P. & Pecerrillo, A. *Bull. volcan.* **51**, 355–378 (1989).
- Métrich, N. & Clochietti, R. *Bull. volcan.* **51**, 185–198 (1989).
- Clochietti, R., Gioncada, A., Mosbach, M. & Sbrana, A. *Acta volcan.* (in the press).
- Francis, P., Oppenheimer, C. & Stevenson, D. *Nature* **366**, 554–557 (1993).
- Rosi, M. *Rc. Soc. Ital. Miner. Petrol.* **36**, 345–368 (1980).
- Tazieff, H. & Tonani, F. *C.r. hebdom. Séanc. Acad. Sci., Paris* **257**, 3985–3987 (1963).
- Chouet, B., Hamisevicz, N. & McGetchin, T. R. *J. geophys. Res.* **79**, 4961–4976 (1974).
- Blackburn, E. A., Wilson, L. & Sparks, R. S. J. *geol. Soc. Lond.* **132**, 429–440 (1976).
- Settle, M. & McGetchin, T. R. *J. Volcan. geotherm. Res.* **8**, 45–58 (1980).
- Capaldi, G. et al. *Bull. volcan.* **41**, 259–285 (1978).
- De Fino, M. et al. *Rc. Soc. Ital. Miner. Petrol.* **43**, 1021–1038 (1988).
- Stoiber, R. E., Malone, G. B. & Bratton, G. P. *Geol. Soc. Am. Abstr. Progr.* **10**, 148 (1978).

15. Stoiber, R. E., Malinconico, L. L. & Williams, S. N. in *Forecasting Volcanic Events* (eds Tazieff, H. & Sabroux, J.-C.) 425–444 (Elsevier, Amsterdam, 1983).
16. Carbonnelle, J. & Zettwoog, P. *Bull. Prog. Interdisciplinaire Rech. Eruptions Volcan. CNRS* **55** (1982).
17. Giberti, G., Jaupart, C. & Sartoris, G. *Bull. volcan.* **54**, 535–541 (1992).
18. Stoiber, R. E., Williams, S. N. & Huebert, B. J. *J. Volcan. geotherm. Res.* **33**, 1–8 (1987).
19. Caltabiano, T. & Romano, R. *Boll. Grup. Naz. Vulcan. ital.* **4**, 133–145 (1988).
20. Allard, P. *et al. Nature* **351**, 387–391 (1991).
21. Bluth, G. J. S., Schnetzler, C. C., Krueger, A. J. & Walter, L. S. *Nature* **366**, 327–329 (1993).
22. Ripepe, M., Rossi, M. & Saccorotti, G. *J. Volcano. geotherm. Res.* **54**, 335–351 (1993).
23. McGetchin, T. R. & Chouet, B. A. *Geophys. Res. Lett.* **6**, 317–320 (1979).
24. Morelli, C. *et al. Boll. Geofis. teor. appl.* **18**, 183–210 (1975).
25. Clocchiatti, R. *Bull. volcan.* **44**, 339–357 (1981).
26. Wallace, P. & Carmichael, I. S. E. *Geochim. cosmochim. Acta* **56**, 1863–1874 (1992).
27. Jaupart, C. & Vergnolle, S. *Nature* **300**, 427–429 (1988).
28. Andres, R. J. *et al. J. Volcan. geotherm. Res.* **46**, 323–329 (1991).
29. Stoiber, R. E. & Jepsen, A. *Science* **182**, 577–578 (1973).
30. Rose, W. I., Stoiber, R. E. & Malinconico, L. L. in *Orogenic Andesites and Related Rocks* (ed. Thorpe, R. S.) 669–676 (Wiley, New York, 1982).
31. Casadevall, T. *et al. Science* **221**, 1383–1385 (1983).
32. Westrich, H. R. & Gerlach, T. M. *Geology* **20**, 867–870 (1992).
33. Stoiber, R. E., Williams, S. N. & Huebert, B. J. *J. geophys. Res.* **91**, 12215–12231 (1986).
34. Allard, P. thesis. Univ. Paris 7 (1986).
35. Ntepe, N. & Dorel, J. J. *J. Volcan. geotherm. Res.* **43**, 235–251 (1990).
36. Killingley, J. S. & Muenow, D. W. *Geochim. Cosmochim. Acta* **39**, 1467–1473 (1975).
37. Gerlach, T. M. *J. geophys. Res.* **91**, 12177–12185 (1986).
38. Chaigneau, M. C. *Reb. Seanc. Acad. Sci., Paris* **261**, 2241–2244 (1965).
39. Martini, M. *Bull. glob. volcan. Network* **18**, 5–21 (1991).
40. Carroll, M. R. & Rutherford, M. J. *J. geophys. Res.* **90**, 601–612 (1985).

ACKNOWLEDGEMENTS. We dedicate this work to the memory of G. B. Malone, who lost his life during the first COSPEC campaign on Stromboli in 1975. Heating stage thermometry on pyroxenes were performed by N.M. in the Vernadsky Institute, Moscow, thanks to A. Sobolev. We are grateful to F. Barberi and M. Dall'Aglia for their support, and to G. Brandeis, T. Caltabiano, G. Giberti, P. Newton, G. Sartoris, S. Tait and S. Vergnolle for useful discussions. S. N. Williams helped us to improve the manuscript. Airborne measurements were supported by the CEA (Commissariat à l'Energie Atomique, France) and the GNV (Gruppo Nazionale per la Vulcanologia, Italy).

Prior experience enhances pattern discrimination in insect vision

S. W. Zhang & M. V. Srinivasan

Centre for Visual Science, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra, ACT 2601, Australia

It is well known that prior knowledge or experience aids us tremendously in uncovering objects that are poorly visible, partially hidden or camouflaged^{1–3}. Is such enhancement in performance unique to higher animals? Here we find that bees cannot be trained directly to distinguish between differently shaped, camouflaged figures. They can, however, learn to break the camouflage and make the discrimination if they are trained initially on a simpler task that exposes them to shapes that are presented later in camouflage. Evidently, even organisms with relatively simple nervous systems can use prior experience to advantage in processing visual images.

Few who view the scene in Fig. 1 for the first time would see a familiar object, especially if they were unaware of the picture's content. Once the camouflaged object has been discovered, however, it is detected and recognized readily every time the picture is re-encountered. Evidently, prior experience or knowledge aids the visual system significantly in the task of uncovering objects^{1–3}. Is the ability to enhance processing in this way restricted to highly developed visual systems, such as those of humans and higher mammals? Here we investigate whether bees are able to use prior experience to advantage in detecting objects and discriminating their shapes.

Using a Y-maze (Fig. 2a), we attempted to train bees to distinguish between two shapes—a ring and a disc—when each shape was presented in a camouflaged fashion as a textured figure 6 cm in front of a similarly textured background (Fig. 2b). We found that bees were unable to learn to make this discrimination, despite lengthy training incorporating over 100 rewards per bee (Fig. 3).

Next, we examined whether bees could learn to distinguish the camouflaged stimuli if they were first trained on a related, but simpler discrimination. To this end, we trained a fresh group of bees to distinguish between a black ring and a black circle, each presented 6 cm in front of a white background. The ring and the circle were of the same size and shape as their textured counterparts, and their spatial configuration in relation to the background was identical to that in the previous experiment. The bees were able to learn this new task: the average choice frequency in favour of the rewarded stimulus (ring) was 63.1%, and this was significantly different from random choice ($P < 0.005$, χ^2 test; Figs 3a, 4a). Then the black figures were replaced by textured figures and the white backgrounds by similarly textured backgrounds, so that the new pair of stimuli was identical to that used in the first experiment (Fig. 2b). The bees now showed a clear ability to distinguish between the camouflaged stimuli (Figs 3b, 4b). The bees that were pre-trained on the black figures performed significantly better than those that were trained directly on the camouflaged stimuli (compare with Fig. 2b).

Similar results were obtained when we used a different pair of camouflaged shapes—textured horizontal and vertical bars—presented 6 cm in front of a similarly textured background. Bees were unable to learn to distinguish between the camouflaged stimuli when they were trained directly on them (Fig. 2c). But they were able to learn to do so when they were pre-trained on black bars presented in front of a white background, using a procedure analogous to that described for Fig. 4a, b. At the end of the pre-training phase, the bees displayed a choice frequency of 73.6% for the correct stimulus (black horizontal bar; $n = 811$, $P < 0.001$; data not shown). At the end of the subsequent training on the textured bars, they displayed a choice frequency of 65.1% for the correct stimulus (textured horizontal bar; $n = 189$, $P < 0.001$; data not shown).

Distinguishing the stimuli in Fig. 2b (or Fig. 2c) is a relatively complex task. It is very unlikely that the bees use stereoscopic vision to detect the camouflaged figures, given their relatively small interocular separation, poor visual acuity⁴ and the small disparity between figure and background. Rather, it appears that they use relative motion cues generated by their own motion in flight to detect the camouflaged figures and discriminate their shapes⁵. The boundary of each figure is signalled by the difference in apparent motion (motion parallax) between the figure and its immediate background. There is good evidence that bees can detect such boundaries by using motion parallax as a cue⁵,



FIG. 1 A familiar, but camouflaged object (readers experiencing difficulty in recognizing the Dalmatian dog may wish to view the picture upside-down). Photo courtesy R. C. James. Reprinted from Lindsay and Norman¹, with permission of authors and publishers.

but until now it has been impossible to train bees to discriminate the shapes of figures whose boundaries are defined solely by motion parallax⁵. Evidently, bees are unable to learn to make this complex discrimination in a single step. However, our results show that bees can learn such a task if they are pre-trained on

a related but simpler discrimination that gives them a 'hint' as to what is expected of them ultimately.

What is the nature of the information that the bees acquire from the step-by-step training? Do they merely acquire information pertaining to the shapes of the figures that are to be discriminated, or do they learn something more general, such as the ability to distinguish between any shapes that are defined solely by motion-parallax cues? To investigate this question, we examined whether the group of bees trained to distinguish between the camouflaged ring and disc (using the step-by-step training procedure shown in Fig. 4*a, b*) could next be trained to distinguish between a different pair of camouflaged stimuli, namely, textured horizontal and vertical bars. We found that the bees were able to learn to make this discrimination (Fig. 4*c*)—albeit after a brief dip in the step-by-step learning curve at the start of training phase *c* in Fig. 3, when they evidently noticed that the camouflaged shapes had changed, and were learning to discriminate the new shapes. This is in contrast to the bees that were trained directly on the textured bars: they could never learn the task (Fig. 2*c*). Thus, the step-by-step training procedure that we have adopted here teaches the bees more than just the specific shapes of the figures to be distinguished; it teaches them to use motion parallax cues to extract the shapes of camouflaged figures in a more general sense. Furthermore, the dip and subsequent recovery in the learning curve when the stimuli are changed from textured ring and circle to textured bars (the transition from phase *b* to phase *c* in Fig. 3) indicate that the bees do not abandon motion parallax as a potential cue during the brief period of 'confusion' (about 10 rewards per bee, or 1 hour).

Is the facility that the bees acquire from the two-step training genuinely a consequence of acquiring relevant prior experience during the first step, or is it merely a nonspecific enhancement of performance produced by pre-training? To examine this question, we investigated whether a group of bees trained to distinguish the black-and-white stimuli of Fig. 4*a* (ring versus disc)

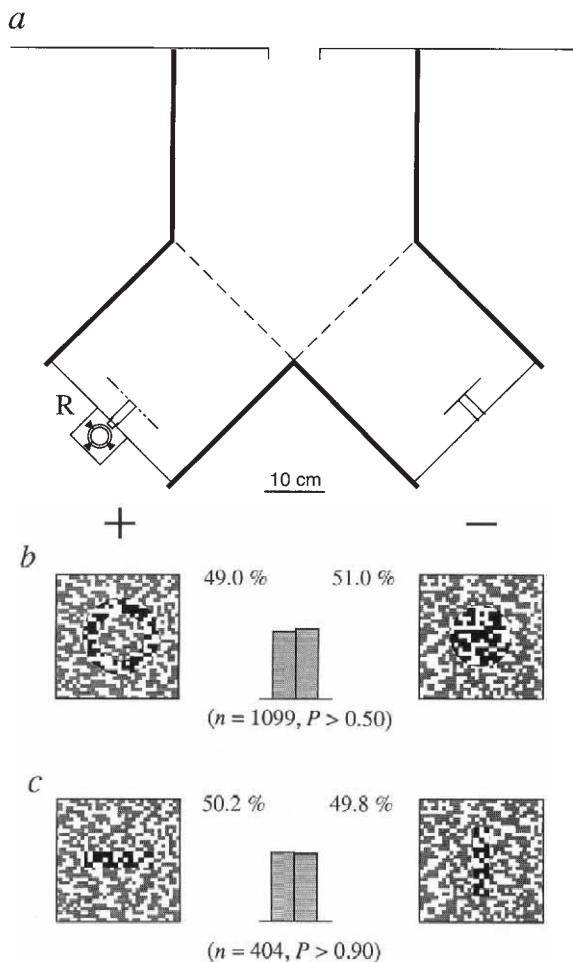


FIG. 2 *a*, Y-maze apparatus for training bees. A group of 10 freely flying bees was marked and trained to enter the apparatus which presented two stimuli, one on the vertical end wall of each tunnel. One stimulus (termed positive, and denoted by +) offered a reward of sugar water, R, placed in the box behind the wall, which the bees could reach through a tube. The other stimulus (termed negative, and denoted by -) carried no reward. The positions of the positive and negative stimuli were interchanged every 10 minutes, and the reward was moved with the positive stimulus during the entire experiment. This prevented the bees from developing a preference for one of the tunnels, and cancelled the effect of any residual side preferences when learning performance was assessed. The bees' learning performance was measured in terms of the percentage of choices that they made in favour of the two stimuli, and statistical tests for significant differences from random choice. A bee's choice was determined by noting which tunnel the bee entered first, when it arrived at the apparatus. Control experiments, using identical stimuli in both tunnels assured us that the bee's choices were not influenced by olfactory cues. For details of the training procedure, see ref. 9. Bees trained in this way were unable to distinguish between *b*, a textured ring and a textured circle, or *c*, between textured horizontal and vertical bars when these figures were presented 6 cm in front of similarly textured backgrounds. For clarity, the textured figures are shown darker than the backgrounds in the illustrations. The percentage figures and the bars depict the relative frequencies of choices in favour of the positive and negative stimuli respectively, as measured during the end of each training phase. *n*, Number of choices analysed; *P*, value associated with a χ^2 test for significant differences from random-choice behaviour.

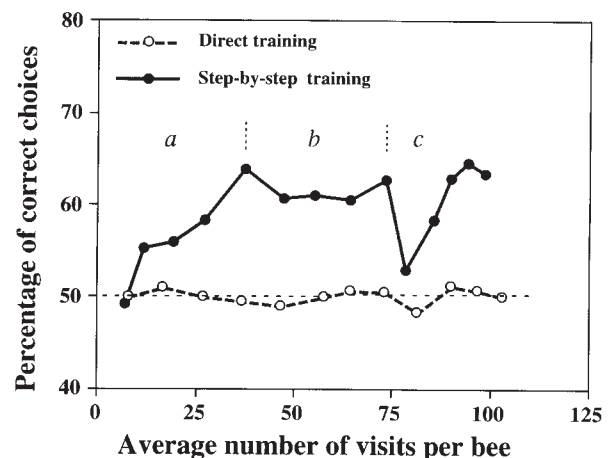


FIG. 3 Learning curves corresponding to the training experiments illustrated in Figs 2*b* and 4. The bees were unable to learn to distinguish the stimuli of Fig. 2*b* when trained directly on them, even after each bee had received over 100 rewards, on average, as illustrated by the dashed curve in the present figure. However, bees were able to learn the same task when they were trained in a step-by-step fashion, as described in Fig. 4 and illustrated by the solid curve here. (The segments labelled *a*, *b* and *c* on the step-by-step training curve correspond to training phases using the stimuli illustrated in Fig. 4*a-c* respectively.) Furthermore, bees trained in this way were also able to learn to distinguish between novel camouflaged shapes, such as textured horizontal and vertical bars (segment *c* on the step-by-step training curve). Each data point on the learning curves depicts the choice frequency in favour of the positive stimulus, as measured by analysing all of the choices made since the previous point.

could learn directly to distinguish the differently shaped, camouflaged stimuli of Fig. 4c (horizontal versus vertical textured bar). In this two-step training, the cues that are germane to the first task are irrelevant to the second. We found that although these bees learned the first task (as before), they were unable to learn the second (choice frequency for correct stimulus, 49.3%; $n = 369$; $P > 0.85$). Thus, pre-training enhances the bees' subsequent performance only when it contains cues that are relevant to the subsequent task. Furthermore, the finding that bees are able to learn to transfer discrimination from the camouflaged ring and disc (Fig. 4b) to the camouflaged horizontal and vertical bar (Fig. 4c) but not from the black/white ring and disc (Fig. 3a) to the camouflaged horizontal and vertical bar (as described above) suggests that the enhancement in discrimination performance is not merely the result of learning to distinguish motion signals of different magnitudes that would arise from scanning the patterns in the horizontal and vertical directions. Rather, it appears that the bees are learning to use motion-parallax cues to extract information on shape.

Our findings do not reveal precisely in what way the pre-training modifies processing of information in the visual pathway. At one extreme, the pre-training could generate an abstract, high-level representation of the objects to be discriminated. This representation could facilitate detection and discrimination of the same objects when they are camouflaged, partially occluded, poorly visible or viewed under different conditions. Such schemes are envisaged in current models of 'top-down' processing in human cognition and computer vision^{1-3,6-8,10}. At the other extreme, the pretraining could simply promote extraction of a number of related, low-level visual cues that are potentially use-

ful to the discrimination, by sensitizing the appropriate visual interneurons. These interneurons could be situated in a pathway that is associated with perception, or in a pathway that simply mediates a behavioural reflex which guides the bee toward the correct stimulus.

Irrespective of these unresolved questions, our results demonstrate clearly that the ability to use prior experience to enhance visual performance is not a faculty that is restricted to primates or higher mammals; it is possessed even by creatures with brains that weigh less than a milligram and carry <0.01% as many neurons as we do. The neural basis of this phenomenon is largely unknown, although it is currently the subject of much interest, speculation and modelling⁶⁻⁸. The relatively simple nervous system of the bee offers a promising substrate in which to explore the underlying mechanisms. □

Received 28 October 1993; accepted 11 January 1994.

1. Lindsay, P. H. & Norman, D. A. *Human Information Processing: An Introduction to Psychology* (Academic, New York, 1977).
2. Goldstein, E. B. *Sensation and Perception* (Wadsworth, Belmont, CA, 1989).
3. Cavanagh, P. in *Representations of Vision* (eds Gorea A., Fregniac, Y., Kapoula, Z. & Findlay, J.) 295-304 (Cambridge Univ. Press, Cambridge, 1991).
4. Srinivasan, M. V., Lehrer, M. & Horridge, G. A. *Proc. R. Soc. Lond. B* **238**, 331-350 (1990).
5. Collett, T. S. & Harkness, L.I.K. in *Analysis of Visual Behaviour* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. J.) 111-176 (MIT Press, Boston, 1982).
6. Li, Z.-N. & Uhr, L. *IEEE Trans. Systems, Man Cybern.* **17**, 250-263 (1987).
7. Colby, C. L. *J. Child Neurol.* **6**, S90-S118 (1991).
8. Mumford, D. *Biol. Cybern.* **66**, 241-251 (1992).
9. van Hateren, H., Srinivasan, M. V. & Wait, P. J. *comp. Physiol. A* **167**, 649-654 (1990).
10. Snyder, A. W. & Barlow, H. B. *Nature* **331**, 117-118 (1988).

ACKNOWLEDGEMENTS. We thank T. Collett, M. Ibbotson, M. Lehrer, E. Maddess and R. Wehner for critical comments on the manuscript.

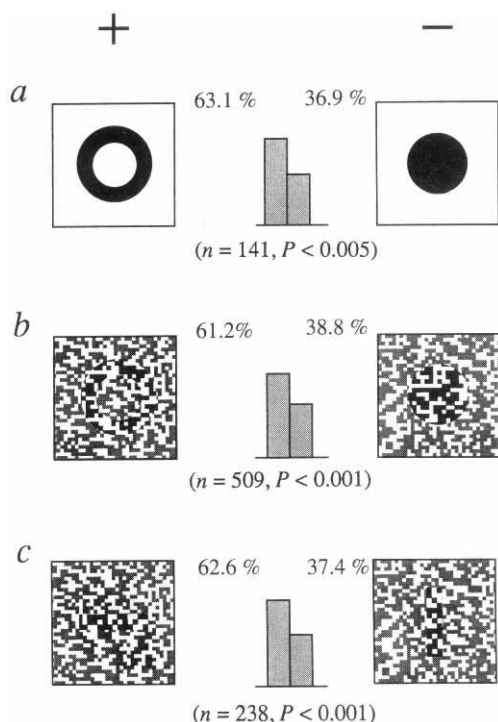


FIG. 4 a, b, Results of a separate experiment, using a fresh group of bees that were trained ultimately to distinguish between the stimuli shown in Fig. 2b, but in two steps. First, the bees were trained to distinguish between a black ring and a black disc, each presented in front of a white background. They learned this task well (a). The same bees were then trained with the textured shapes presented in front of the textured background. These bees learned to discriminate the camouflaged shapes well (b), in contrast to the bees that were trained directly on the camouflaged shapes (compare with Fig. 2b). Furthermore, the group of bees trained in the step-by-step fashion could also learn to discriminate other camouflaged shapes (c).

Unusual permeability properties of gastric gland cells

Steven J. Walsbren*†, John P. Geibel*†, Irvin M. Modlin† & Walter F. Boron*

* Department of Cellular and Molecular Physiology and † Gastro-intestinal Surgical Pathobiology Research Unit, Department of Surgery, Yale University School of Medicine, New Haven, Connecticut 06510, USA

PHYSIOLOGISTS have long pondered the riddle of why the stomach is itself not digested by the very juice it secretes. One explanation is that a mucus-bicarbonate barrier, coating the stomach lumen as well as superficial portions of gastric glands, prevents autodigestion¹. However, this leaves unanswered the question of what protects cells deeper in the glands, which seem to lack a mucus barrier². These are the parietal and chief cells, which secrete acid and pepsin. Using perfused single gastric glands from rabbit, we recently found that intracellular pH is uniquely resistant to extreme degrees of luminal acidification², suggesting that the apical (luminal) barrier might also exclude ammonia and carbon dioxide, to which cell membranes are generally highly permeable^{3,4}. We now show that this is indeed the case. There are three reports of membranes with very low permeabilities to NH₃ (refs 5-7), and none of membranes impermeable to CO₂.

We used intracellular pH (pH_i) changes to assess the permeability of apical and basolateral membranes of parietal cells and chief cells to NH₃ and CO₂. Perfusing single gastric glands, we loaded cells with the pH-sensitive dye BCECF, determining the pH_i of many individual cells using digital processing of fluorescent images². We found that the apical barriers of both parietal and chief cells are impermeable to NH₃ and CO₂, as well as to NH₄⁺ and HCO₃⁻.

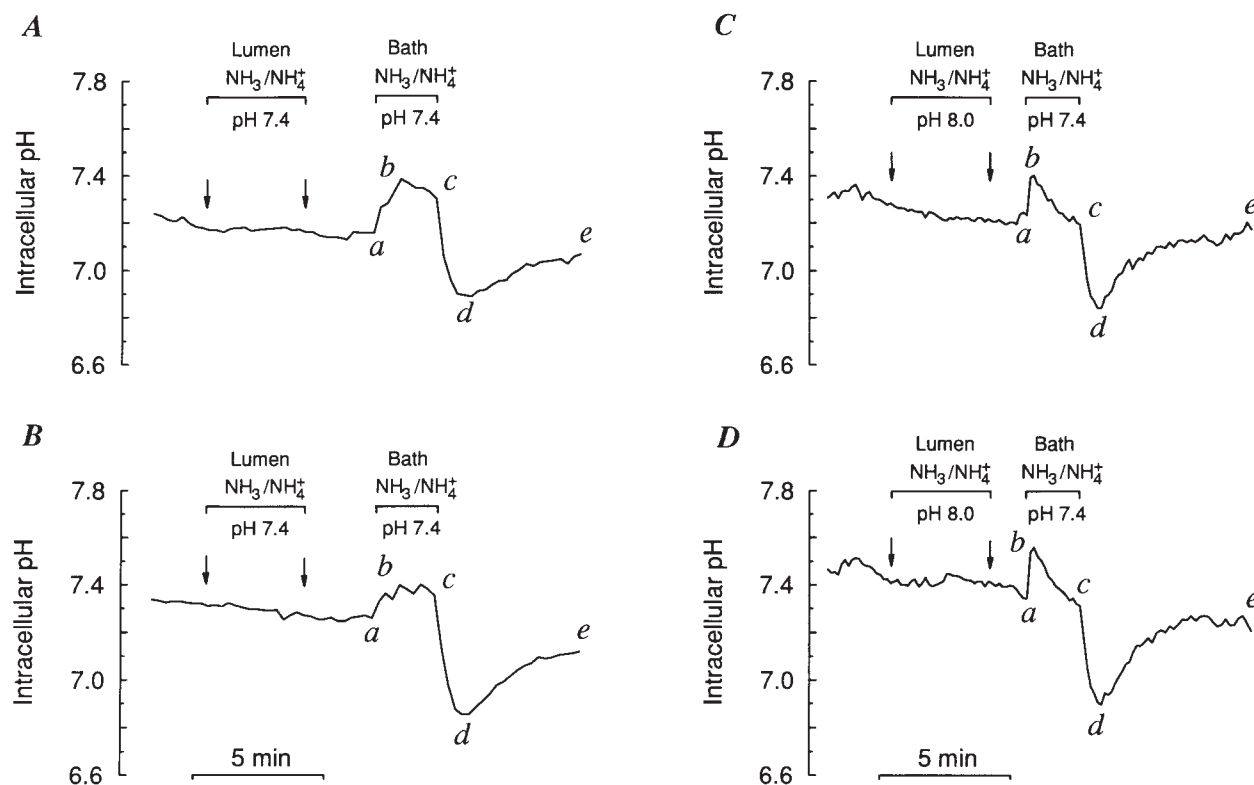


FIG. 1 Effect of $\text{NH}_3/\text{NH}_4^+$ on the intracellular pH of parietal and chief cells in isolated, perfused gastric glands. A, Response of a single parietal cell to application and removal of a solution at pH 7.4 containing 20 mM $\text{NH}_3/\text{NH}_4^+$, first from lumen and then from bath. For 62 cells from 10 glands taken from 8 animals ($n=62/10/8$), the mean pH_i changes for bath $\text{NH}_3/\text{NH}_4^+$ pulses were $+0.33 \pm 0.03$ in segment *ab* (significantly different from zero; $P < 0.001$) and -0.47 ± 0.03 ($P < 0.001$) in *cd*. The comparable figures for luminal $\text{NH}_3/\text{NH}_4^+$ pulses were -0.003 ± 0.003 (not significantly different from zero; NS) and -0.003 ± 0.002 (NS), respectively. B, Response of a single chief cell from same experiment as in A. Mean pH_i changes ($n=45/7/6$): bath $\text{NH}_3/\text{NH}_4^+$ pulses: $+0.28 \pm 0.02$ ($P < 0.001$) (*ab*), -0.45 ± 0.02 ($P < 0.001$) (*cd*); luminal pulses: -0.002 ± 0.003 (NS) and -0.002 ± 0.001 (NS). C, Response of a single parietal cell, first to luminal application and removal of a pH 8.0 solution containing 20 mM $\text{NH}_3/\text{NH}_4^+$, and then to basolateral application and removal of a pH 7.4 solution containing 20 mM $\text{NH}_3/\text{NH}_4^+$. Mean pH_i changes ($n=4/2/1$): bath pH 7.4 $\text{NH}_3/\text{NH}_4^+$ pulses: $+0.22 \pm 0.03$ ($P < 0.001$) (*ab*), -0.38 ± 0.07 ($P < 0.001$) (*cd*); luminal pH 8.0 pulses: 0.005 ± 0.003 (NS) and -0.005 ± 0.005 (NS). D, Response of a single chief cell from the same experiment as in C. Mean pH_i changes ($n=$

11/2/1): bath $\text{NH}_3/\text{NH}_4^+$ pulses: $+0.21 \pm 0.02$ ($P < 0.001$) (*ab*), -0.40 ± 0.06 ($P < 0.001$) (*cd*); luminal pulses: 0.003 ± 0.003 (NS) and 0.001 ± 0.003 (NS).

METHODS. We describe elsewhere the perfusion of isolated gastric glands, an extension of an approach originally used to perfuse renal tubules¹⁸, as well as the measurement of pH_i using digital processing of fluorescent images². Briefly, rabbit gastric glands were dissected out by hand, transferred to a chamber on the stage of an inverted fluorescence microscope, perfused, and loaded with BCECF. Lumen and bath solutions flowed continuously through CO_2 -impermeable tubing. The order of perturbations (lumen compared with bath) was randomized. Luminal perfusion was confirmed, before and after each experiment, by perfusing the lumen with trypan blue. Intracellular dye was alternately excited at 440 nm and 490 nm, while emissions (I_{490} and I_{440}) were monitored at 530 nm using an image intensifier and a CCD-type television camera. A digital image processor was used to compute I_{490}/I_{440} . After each experiment, we did a single-point dye calibration at pH_i 7.00, using a variant of the high- K^+ /nigericin technique^{19,20}. Bath temperature was 37 °C. Statistical significance was determined using a one-tailed Student's *t*-test.

When a gastric gland, from either lumen or bath (that is, the basolateral surface), was exposed to a solution at pH 7.4 containing 20 mM $\text{NH}_3/\text{NH}_4^+$, the addition of $\text{NH}_3/\text{NH}_4^+$ to the lumen had no effect on the pH_i of either a parietal (Fig. 1A) or a chief (Fig. 1B) cell. However, introducing the same solution via the bath caused the expected pH_i changes^{8,9}: a rapid increase due to NH_3 influx (Fig. 1A, B; segment *ab*), followed by a slower decrease (*bc*) due to NH_4^+ uptake and other acidifying processes. Removal of the $\text{NH}_3/\text{NH}_4^+$ pulses caused a rapid decrease in pH_i (Fig. 1A, B; segment *cd*) due to NH_3 efflux, followed by a slower recovery (*de*), presumably due to active transport of acid out of the cell. In a broader series of experiments, luminal $\text{NH}_3/\text{NH}_4^+$ pulses lasting as long as 6 min never produced a discernible change in pH_i parietal cells ($n=62$ cells) or chief cells ($n=45$), whereas bath $\text{NH}_3/\text{NH}_4^+$ pulses produced consistently large pH_i changes. In two other experiments (not shown), we first removed Na^+ (replaced with *N*-methyl-D-glucammonium) from the luminal solution, and then switched to a Na^+ -free solution containing

135 mM $\text{NH}_3/\text{NH}_4^+$. Neither luminal solution change affected pH_i .

We can rule out two explanations for the failure of luminal NH_3 to alkalinize cells. First, luminal acid secretion may titrate all NH_3 to NH_4^+ . However, inhibiting acid secretion with the H_2 -receptor blocker cimetidine (100 μM ; $n=3$), or with both cimetidine and the H^+ - K^+ pump inhibitor omeprazole (10 μM ; $n=3$), did not alter the results. Second, apical permeability to NH_4^+ may be so high that the acidifying effects of NH_4^+ influx into the parietal and chief cells exactly balance the alkalinizing effects of NH_3 influx. However, the pH_i was unaffected when the pH of the luminal 20 mM $\text{NH}_3/\text{NH}_4^+$ solution was increased from 7.4 to 8.0, thereby increasing luminal $[\text{NH}_3]$ from 0.40 to 1.46 mM without substantially decreasing the NH_4^+ gradient across the apical membrane (Fig. 1C, D).

When a gland, from either lumen or bath, was exposed to a solution at pH 7.4 containing 5% $\text{CO}_2/22$ mM HCO_3^- , both a parietal cell (Fig. 2A) and a chief cell (Fig. 2B) showed a transi-

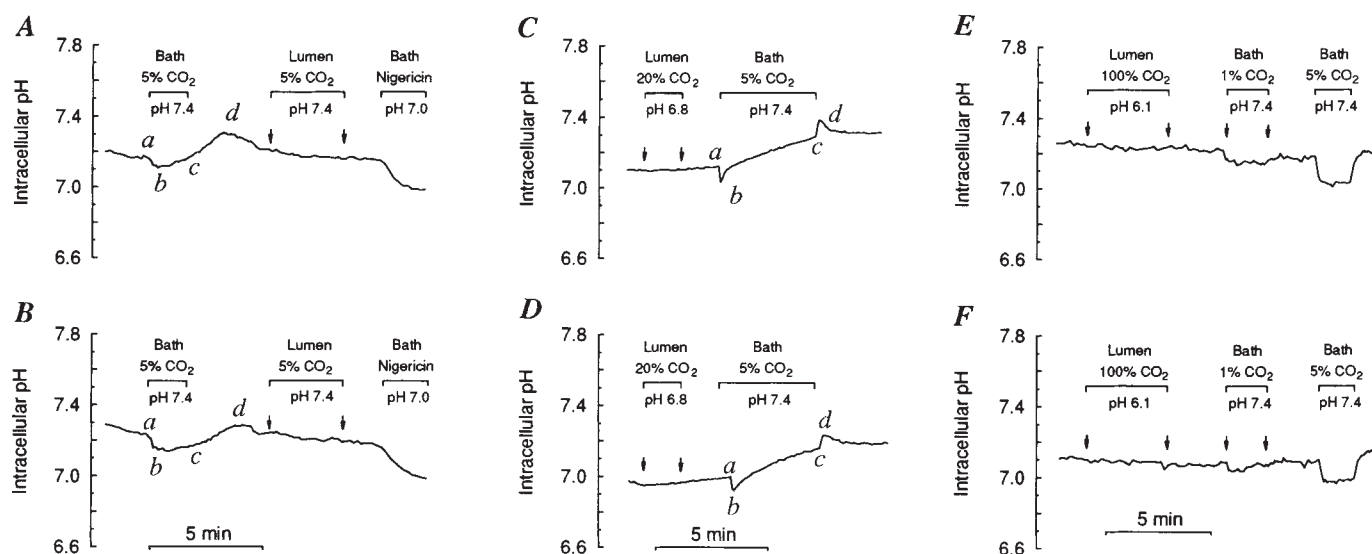


FIG. 2 Effect of $\text{CO}_2/\text{HCO}_3^-$ on intracellular pH of parietal and chief cells in isolated, perfused gastric glands. **A**, Response of a single parietal cell to the application and removal of a solution at pH 7.4 containing 5% $\text{CO}_2/22 \text{ mM HCO}_3^-$, first from bath and then from lumen. The experiment was ended with calibration of intracellular BCECF, using a high- K^+ nigericin solution at pH 7.00. Mean pH_i changes ($n=29/4/3$): bath $\text{CO}_2/\text{HCO}_3^-$ pulses: -0.038 ± 0.009 (ab), $+0.117 \pm 0.016$ (cd), $+0.156 \pm 0.021$ (bd) (all results, $P < 0.001$); luminal pulses: $+0.001 \pm 0.001$, -0.000 ± 0.002 and 0.004 ± 0.003 (all results, NS). **B**, Response of a single chief cell from same experiment as in **A**. In other experiments, the segment-bc pH_i increase never waned during exposures lasting as long as 6 min. Mean pH_i changes ($n=25/3/2$): bath $\text{CO}_2/\text{HCO}_3^-$ pulses: -0.047 ± 0.009 ($P < 0.001$) (ab), $+0.076 \pm 0.036$ (cd; significantly different from zero, $P < 0.025$), $+0.184 \pm 0.021$ ($P < 0.001$) (bd); luminal pulses: 0.000 ± 0.001 , -0.004 ± 0.003 and 0.003 ± 0.003 (all results, NS). **C**, Response of a single parietal cell, first to luminal application and removal of a pH 6.8 solution containing 20% $\text{CO}_2/22 \text{ mM HCO}_3^-$, and then to basolateral application and removal of a pH 7.4 solution containing 5% $\text{CO}_2/$

22 mM HCO_3^- . Mean pH_i changes: bath 5% CO_2 pulses ($n=20/4/2$): -0.031 ± 0.007 (ab), $+0.191 \pm 0.016$ (cd), $+0.281 \pm 0.026$ (bd) (all results, $P < 0.001$); luminal 20% pulses ($n=26/4/2$): -0.001 ± 0.005 , 0.003 ± 0.002 and 0.002 ± 0.004 (all results, NS). **D**, Response of a single chief cell from same experiment as in **C**. Mean pH_i changes ($n=12/2/1$): bath 5% CO_2 pulses: -0.096 ± 0.046 ($P < 0.025$) (ab), $+0.226 \pm 0.012$ ($P < 0.001$) (cd), $+0.298 \pm 0.017$ ($P < 0.001$) (bd); luminal 20% pulses ($n=26/4/2$): $+0.003 \pm 0.005$, 0.000 ± 0.004 and 0.000 ± 0.004 (all results, NS). **E**, Response of a single parietal cell, exposed to basolateral 200 μM DIDS, to luminal application and removal of a pH 6.1 solution containing 100% $\text{CO}_2/22 \text{ mM HCO}_3^-$ and to basolateral application of pH 7.4 solutions containing 1% $\text{CO}_2/4.4 \text{ mM HCO}_3^-$ and 5% $\text{CO}_2/22 \text{ mM HCO}_3^-$. **F**, Response of a single chief cell from the same experiment as in **E**. The bath was perfused continuously with the standard air-equilibrated HEPES-buffered solution. As it left the perfusion pipette, a sample of the 100% CO_2 solution was collected under oil (equilibrated with 100% CO_2). The pH of this solution, measured with a miniature, glass pH microelectrode was 6.1.

ent decrease in pH_i (segment ab), due to CO_2 influx, followed by sustained pH_i increases (bc), presumably due to the active uptake of HCO_3^- (ref. 10). These effects were reversed by removing $\text{CO}_2/\text{HCO}_3^-$ (Fig. 2C, D; cd). However, introducing and then removing $\text{CO}_2/\text{HCO}_3^-$ via the lumen had no effect on pH_i in either cell. In a broader series of experiments, introducing 5% $\text{CO}_2/22 \text{ mM HCO}_3^-$ to the lumen for as long as 5 min produced no pH_i changes in either parietal ($n=29$) or chief ($n=25$) cells, whereas these same glands responded to bath $\text{CO}_2/\text{HCO}_3^-$.

By analogy with the $\text{NH}_3/\text{NH}_4^+$ experiments, one explanation for the failure of luminal CO_2 to acidify cells is that apical permeability to HCO_3^- is so high that the alkalinizing effect of HCO_3^- influx into parietal and chief cells exactly nullifies the effect of CO_2 influx. However, exposing the lumen to 20% $\text{CO}_2/22 \text{ mM HCO}_3^-$, rather than 5%/22 mM, failed to decrease pH_i (Fig. 2C, D), even though the gradient for apical CO_2 entry was increased fourfold.

A second explanation for the failure of CO_2 or NH_3 to alter pH_i is that the surface area is far less at the apical than at the basolateral barrier. Early work suggests that the apical:basolateral area ratio is $\sim 3:1$ in murine parietal cells¹¹ and $\sim 1:4$ in canine parietal cells¹². We addressed this issue by perfusing the lumen with 100% CO_2 , or perfusing the bath with the lowest CO_2 concentration producing an easily detectable pH_i decrease. To improve detection of CO_2 -induced acidifications, we blocked HCO_3^- transport by adding DIDS (4,4'-diisothiocyanato-2,2'-stilbene disulphonate) to the bath. Figure 2E and F show the pH_i responses of a parietal and a chief cell from the same gland to luminal 100% CO_2 (22 mM HCO_3^- ; pH 6.1) as well as to basolateral 1% CO_2 (4.4 mM HCO_3^- ; pH 7.4) and 5% CO_2

(22 mM HCO_3^- ; pH 7.4). Although luminal 100% CO_2 produced no discernible pH_i changes, basolateral 1% CO_2 elicited a small but reproducible decrease in pH_i ; 5% CO_2 produced an even greater response. In 41 parietal cells (8 glands) and 26 chief cells (8 glands), luminal 100% CO_2 caused no significant decrease in pH_i , which averaged 0.005 ± 0.005 and 0.018 ± 0.013 (mean $\pm$ s.e.m.), respectively. However, 1% CO_2 in the bath caused a significant decrease in pH_i in both parietal (0.08 ± 0.01 ; $P < 0.001$) and chief (0.06 ± 0.01 ; $P < 0.001$) cells. Extending to epithelia cells an earlier model⁸ describing the effects of $\text{CO}_2/\text{HCO}_3^-$ fluxes on pH_i , we found that our parietal cell data for 100% luminal and 1% basolateral CO_2 require that the basolateral permeability-area product be more than 10^3 -fold greater than the apical product. Thus, differences in surface area, by themselves, probably cannot account for the failure of luminal CO_2 or NH_3 to alter pH_i .

A third explanation for the failure of apical $\text{CO}_2/\text{HCO}_3^-$ (or $\text{NH}_3/\text{NH}_4^+$) to elicit a change in intracellular pH is that apical permeability to CO_2 (NH_3) may be so high that virtually all $\text{CO}_2/\text{HCO}_3^-$ ($\text{NH}_3/\text{NH}_4^+$) disappears from the lumen before reaching the cells where pH_i is measured. However, we found that pH_i measurements from cells immediately adjacent to the perfusion pipette did not differ from measurements taken from cells farther away. Further evidence comes from experiments (not shown) in which we perfused the lumen with unbuffered solutions containing the pH-sensitive dye hydroxypyrenetrissulphonate (HPTS)². In five experiments, switching the luminal perfusate from 5% $\text{CO}_2/\text{pH 7.4}$ to 100% $\text{CO}_2/\text{pH 6.1}$ caused a large, abrupt, sustained and uniform decrease in luminal pH (pH_l) along the entire gland. Had CO_2 left the lumen, pH_i would

have become increasingly more alkaline at increasing distances from the perfusion pipette. We confirmed that we could easily detect a pH_i change from 6.1 to 6.4. In five other experiments, the addition of 80 mM $\text{NH}_3/\text{NH}_4^+$ to a pH 7.4 perfusate did not alter pH_i , which remained 7.40 along the entire gland. Had NH_3 left the lumen, pH_i would have decreased, and the acidification would have increased with distance from the perfusion pipette. We verified that we could detect pH_i changing from 7.4 to 7.1. Thus, it appears that the $\text{CO}_2/\text{HCO}_3^-$ and $\text{NH}_3/\text{NH}_4^+$ delivered via the perfusion pipette actually reached the apical barriers of the parietal and chief cells. Additional evidence that the luminal perfusate can affect pH_i comes from the observation that the recovery of pH_i from acute loads in parietal cells is enhanced by raising luminal $[\text{K}^+]$, but inhibited by introducing luminal SCH28080 (ref. 13).

Our data indicate that, although the basolateral barriers of parietal and chief cells have normal permeability characteristics for $\text{NH}_3/\text{NH}_4^+$ and $\text{CO}_2/\text{HCO}_3^-$, the apical barriers have no detectable permeability to NH_3 , NH_4^+ , CO_2 or HCO_3^- . Impermeability to NH_4^+ distinguishes the parietal and chief cell apical barriers from that of the renal thick ascending limb⁶ as well as the *Xenopus* oocyte plasma membrane⁷, which have very low NH_3 permeabilities but substantial pathways for NH_4^+ . It is believed that NH_3 (ref. 14) and CO_2 (ref. 15) traverse cell membranes via lipid pathways. Kikeri *et al.* speculated that the low NH_3 permeability of the apical barrier of the renal thick ascending limb is due to an unusual lipid composition⁶. Indeed, the NH_3 permeability of membrane vesicles derived primarily from porcine parietal cell apical membranes is about one-tenth that

of lecithin unilamellar vesicles¹⁶. Swimbladders of some fish are lined with guanine crystals that may impede oxygen diffusion¹⁷, although such crystals have not been identified in gastric glands or renal tubules. Regardless of the mechanism, the unusual permeability properties of parietal and chief cell apical barriers may have evolved to protect these cells from the harsh environment of the gastric-gland lumen, which contains high levels of acid, the proteolytic enzyme pepsin and, perhaps, noxious ingested agents. □

Received 27 October 1993; accepted 27 January 1994.

1. Silen, W. *Am. J. Med.* **79**, 45–48 (1985).
2. Waisbren, S. J., Geibel, J. P., Boron, W. F. & Modlin, I. M. *Am. J. Physiol.* (in the press).
3. Jacobs, M. H. *J. gen. Physiol.* **5**, 181–188 (1922).
4. Jacobs, M. H. *Am. J. Physiol.* **51**, 321–331 (1920).
5. Garvin, J. L., Burg, M. B. & Knepper, M. A. *Am. J. Physiol.* **255**, F57–F65 (1988).
6. Kikeri, D., Sun, A., Zeidel, M. L. & Hebert, S. C. *Nature* **339**, 478–480 (1989).
7. Burckhardt, B.-C. & Fromter, E. *Pflügers Arch.* **420**, 83–86 (1992).
8. Boron, W. F. & De Weer, P. *J. gen. Physiol.* **67**, 91–112 (1976).
9. Boron, W. F. & De Weer, P. *Nature* **259**, 240–241 (1976).
10. Townsley, M. C. & Machen, T. E. *Am. J. Physiol.* **257**, G350–G356 (1989).
11. Schofield, G. C., Ito, S. & Bolender, R. P. *J. Anat.* **128**, 669–692 (1979).
12. Helander, H. F. & Hirschowitz, B. I. *Gastro* **63**, 951–961 (1972).
13. Geibel, J., Sachs, G. & Modlin, I. *Gastroenterology* **104**, A86 (1993).
14. Golchini, K. & Kurtz, I. *Am. J. Physiol.* **255**, F135–F141 (1988).
15. Gutknecht, J., Bisson, M. A. & Tosteson, F. C. *J. gen. Physiol.* **69**, 779–794 (1977).
16. Priver, N. A., Rabon, E. C. & Zeidel, M. L. *Biochemistry* **32**, 2459–2468 (1993).
17. Lapennas, G. N. & Schmidt-Nielsen, K. *J. exp. Biol.* **67**, 175–196 (1977).
18. Burg, M. B. *Kidney Int.* **22**, 417–424 (1982).
19. Thomas, J. A., Buchsbaum, R. N., Zimniak, A. & Racker, E. *Biochemistry* **81**, 2210–2218 (1979).
20. Boyarsky, G., Ganz, M. B., Sterzel, B. & Boron, W. F. *Am. J. Physiol.* **255**, C844–C856 (1988).

ACKNOWLEDGEMENTS. This work was supported by the NIH.

NMDA-receptor channel diversity in the developing cerebellum

Mark Farrant, Dirk Feldmeyer,
Tomoyuki Takahashi* & Stuart G. Cull-Candy

Department of Pharmacology, University College London,
Gower Street, London WC1E 6BT, UK

* Department of Neurophysiology, Institute of Brain Research,
Faculty of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113,
Japan

In the cerebellum, NMDA (*N*-methyl-D-aspartate) receptors play an important role in neuronal differentiation^{1,2} and excitatory synaptic transmission^{3–5}. During early cerebellar development, marked changes occur in the distribution of messenger RNAs encoding various NMDA-receptor subunits⁶. To determine whether these changes result in the appearance of functionally distinct NMDA receptors^{7–9}, we have recorded single-channel currents in rat cerebellar granule cells during the period of their migration from the external germinal layer to the inner granular layer¹⁰. Here we show that before synapse formation^{10,11}, pre-migratory and migrating granule cells express NMDA receptors possessing single-channel properties similar to those previously described for many central neurons^{12–16}. In contrast, mature post-migratory cells also express an atypical form of NMDA receptor that has a lower single-channel conductance and distinct kinetic behaviour. The properties of these 'low-conductance' channels correspond to those described¹⁷ for recombinant NMDA receptors formed by co-expression of NR1 and NR2C subunits^{8,9}. The NR2C subunit appears postnatally and is found predominantly in the adult cerebellum^{6–8}. Our data demonstrate developmental changes in NMDA-receptor properties at the single-channel level, and suggest that in the cerebellum the expression of a specific subunit protein results in a distinct form of native receptor.

Cerebellar granule cells at various developmental stages (Fig. 1) were studied using patch-clamp methods¹⁸. Whole-cell recordings were made from pre-migratory granule cells in the external germinal layer (EGL), migrating cells in the molecular layer and post-migratory cells in the inner granular layer (IGL) of cerebellar slices from rats at postnatal days 7–14 (P7–P14). Cells were identified by their position within the slice, their morphology and their electrical characteristics (Fig. 1 legend). As shown in Fig. 1*b* and *c*, application of NMDA (50 μM) evoked currents in all migrating and internal cells ($n=14$ and 31) and in 60% of external cells ($n=9$ out of 15). The responses from the different cell types were blocked to a similar extent both by the NMDA antagonist D-2-amino-5-phosphonopentanoic acid (AP5) and at negative potentials by the NMDA-channel blocking ion Mg^{2+} (refs 12, 19) (Fig. 1*d*, *e*). The threefold increase in amplitude of the NMDA response during the period of migration (Fig. 1*c*) was followed by a marked decline in older animals (Fig. 1 legend; see also ref. 20), suggesting that this increase in the density of functional receptors is transient.

The inhibition of granule cell migration by NMDA antagonists has led to the suggestion that NMDA receptors regulate this process¹. Our results provide direct evidence for the expression of functional NMDA receptors by granule cells during this migratory period. In the absence of added NMDA or glutamate, most cells exhibited a background level of NMDA-receptor activation, evident as single-channel openings in the whole-cell record (data not shown). Such tonic activation of NMDA receptors^{21,22} could mediate a non-synaptic trophic action of endogenous glutamate¹, although in slices the possibility of non-specific glutamate leakage from damaged cells cannot easily be excluded.

To investigate further the properties of NMDA receptors at various stages of granule cell development, we examined single-channel currents in outside-out membrane patches. Currents were activated by glutamate or NMDA in virtually all patches from migrating cells (7 out of 7) and internal cells (97 out of 100), and in most patches from external cells (28 out of 38). Figure 2*a* shows typical single-channel events from a cell in the

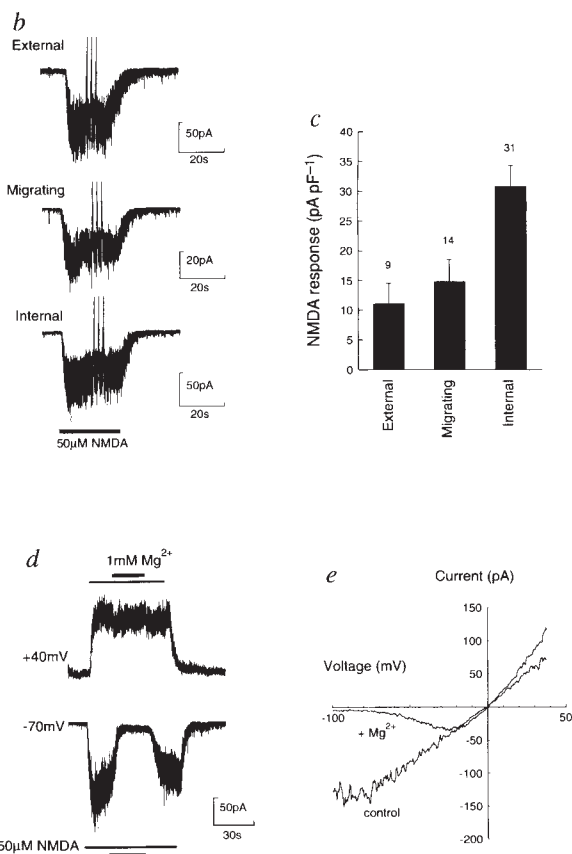
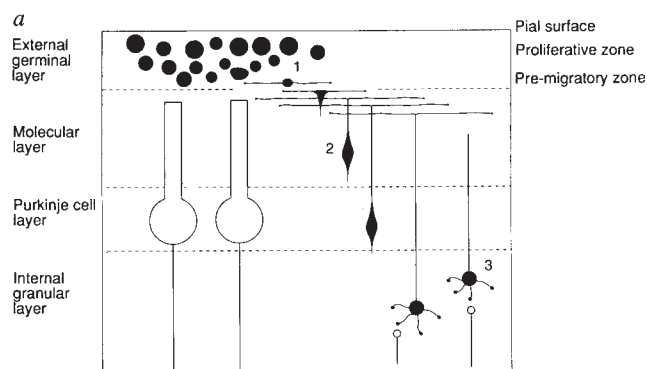


FIG. 1 Whole-cell NMDA responses in developing cerebellar granule cells. **a**, Diagrammatic representation of granule cell migration, showing the position and morphology of granule cells at various stages of development. Recordings from the external germinal layer were obtained mainly from cells in the pre-migratory zone (1), where they either lacked processes or had one or two processes oriented parallel to the pial surface, indicating the extension of parallel fibre axons before cell migration. Migrating granule cells (2) were located at the inner edge of the pre-migratory zone or within the molecular layer and were frequently elongated in the direction of migration, with an apparent leading and/or trailing process. Cells within the internal granular layer (3) possessed three or four short dendrites and a thin axon. **b**, Representative whole-cell responses to NMDA (50 μ M), recorded at -70 mV in the presence of glycine²⁷ (3 μ M), strychnine (200 nM) and tetrodotoxin (300 nM), from an 'external', a 'migrating' and an 'internal' granule cell in a slice from a P11 rat. The bar indicates the duration of NMDA application. The large current deflections are due to ramp changes in the holding voltage, used to generate current-voltage relationships. **c**, Histogram of pooled data of peak current amplitudes normalized to cell capacitance from the three cell types (P7 to P14) (vertical error bars represent s.e.m.); 9 of 15 external cells responded to NMDA, and only responsive cells are included in the mean. The response of internal granule cells was significantly greater than that of either migrating or external cells ($P < 0.01$, unpaired Student's *t*-test). In internal cells of older animals, the NMDA response was markedly reduced (from 30.7 ± 3.7 pA pF⁻¹ ($n = 31$) at P7–P14 to 6.2 ± 0.7 pA pF⁻¹ ($n = 12$) at P22; $P < 0.01$, unpaired Student's *t*-test). **d**, Representative whole-cell NMDA currents from an internal granule cell (P7) showing block by 1 mM Mg²⁺ at -70 mV, but not $+40$ mV. A similar degree of block by Mg²⁺ was observed in granule cells at all three stages of development. For external, migrating and internal cells (P7–P14), the block by 1 mM Mg²⁺ at -70 mV was $86.6 \pm 6.8\%$ ($n = 3$), $94.7 \pm 2.2\%$ ($n = 6$) and $92.8 \pm 1.6\%$ ($n = 7$), respectively (responses to NMDA in older cells (P22) were similarly blocked by Mg²⁺). Likewise, AP5 produced an equivalent block of the responses in the three cell types; corresponding values for 50 μ M AP5 were $94.4 \pm 2.0\%$ ($n = 3$), $89.1 \pm 4.3\%$ ($n = 5$) and $91.2 \pm 1.3\%$ ($n = 7$). These values are not significantly different. **e**, Current-voltage relationship of the response of an internal cell (P7) to 50 μ M NMDA obtained in the absence and the presence of 1 mM Mg²⁺. Similar curves were obtained for cells of the EGL and IGL; in all cases the currents reversed close to 0 mV and showed a region of negative slope conductance between -40 and -100 mV in the presence of Mg²⁺.

METHODS. Parasagittal slices (100–200 μ m) were cut from the cerebella of Sprague-Dawley rats and maintained in a bathing solution containing (mM): NaCl, 125; KCl, 2.5; CaCl₂, 1; NaHCO₃, 26; NaH₂PO₄, 1.25; glucose, 15 or 25; pH 7.4 when bubbled with 95% O₂ and 5% CO₂. Cells on the surface of the slice were viewed with Nomarski differential contrast optics (Zeiss; $\times 320$ – $1,000$). In slices from young animals,

cleaning of the cells, to facilitate seal formation²⁸, was performed occasionally but was not routinely required. In slices from older animals (beyond P16), cleaning proved beneficial, although the overall percentage of successful recordings was low. Patch-pipettes were pulled from thick-walled glass (Clark Electromedical GC150F-7.5), coated with Sylgard resin (Dow Corning 184) and fire-polished to a final resistance of 5–10 M Ω for whole-cell recording or 10–15 M Ω for patch recordings. The pipette (intracellular) solution contained (mM): CsF, 110; CsCl, 30; NaCl, 4; CaCl₂, 0.5; HEPES, 10; EGTA, 5 (adjusted to pH 7.3 with CsOH). In some experiments Lucifer yellow (Li⁺ salt, 0.1%; Sigma) was added to this solution, and filled cells were viewed with fluorescence illumination. Whole-cell recordings were made using an L/M-EPC 7 (List) amplifier. Cell capacitance (external cells, 2.4 ± 0.1 pF, $n = 57$; migrating cells, 3.4 ± 0.3 pF, $n = 25$; internal cells, 3.2 ± 0.3 pF, $n = 33$) was measured from the amplifier and from transient currents produced by 10-mV hyperpolarizing voltage steps, digitized at 100–125 kHz after filtering at 20 kHz (-3 dB Bessel throughout). Drugs were applied by local perfusion from a plastic tube (100 μ m diameter) placed 200 μ m from the recorded cell. Recordings were obtained at 22–25 $^{\circ}$ C. Currents were recorded either on FM tape (Racal Store 4; bandwidth DC to 2.5–5.0 kHz) or on video tape with pulse code modulation (NF Electronic Instruments, RP-880). Current-voltage relationships were obtained by changing the membrane voltage at a rate of 200 mV s⁻¹ from -100 to $+40$ mV (holding potential -70 mV). Curves were produced after subtracting the responses obtained in control solution or in Mg²⁺ (average of 6–10 ramps obtained before and after NMDA application) from that obtained during the response to NMDA or NMDA plus Mg²⁺ (3–5 ramps). All values in the text are expressed as mean $\pm$ s.e.m.

external layer. In the presence of 1 mM Ca²⁺ the main conductance levels were approximately 50 and 40 pS. The corresponding amplitude histogram, indicating the relative frequencies of the two conductance levels, shows that most of the openings (84%) were to the '50-pS' state. We observed similar single-channel openings in all patches from migrating cells (not shown) and most patches from early post-migratory cells (Fig. 2b). In each case, the currents were indistinguishable from those recorded in the EGL, both in the main conductance levels present (50 and

40 pS) and in their relative frequencies. An additional subconductance level of about 30 pS was detected in a few patches from cells of the IGL, but when present these accounted for only 5% of the openings. Direct transitions were apparent between the 50-, 40- and 30-pS levels, indicating that all three conductance states arose from a single channel type¹⁵.

In contrast to the 50/40-pS events present during the early stages of granule cell development, NMDA-receptor channels with novel characteristics were observed in patches from internal

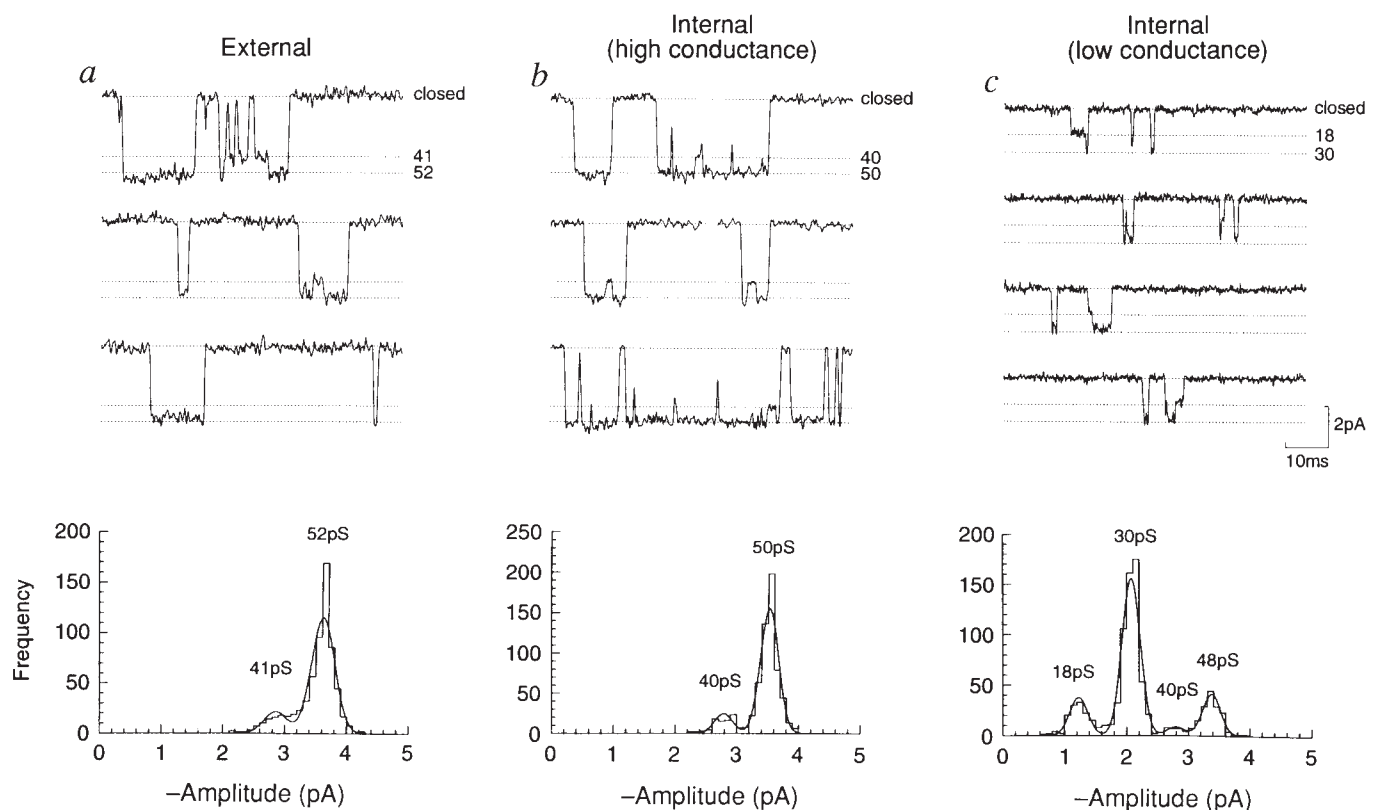


FIG. 2 Single-channel currents activated by NMDA from pre- and post-migratory granule cells. *a*, Currents activated by 10 μ M NMDA (3 μ M glycine) in an outside-out patch from a pre-migratory cell in the EGL (-70 mV). Openings are to a main conductance state of 52 pS and subconductance of 41 pS, with clear transitions between levels. Dashed lines indicate the mean current amplitudes derived from the corresponding single-channel amplitude distribution (histogram below) which illustrates the relative frequencies of openings to the different levels (52 pS, 84.4%; 41 pS, 15.6%). For seven patches from external cells, mean conductances of 49.8 ± 0.4 pS ($85 \pm 1\%$) and 39.9 ± 0.4 pS ($15 \pm 1\%$) were obtained. For migrating cells ($n=3$), the corresponding values were 47.2 ± 2.7 pS ($82 \pm 3\%$) and 37.4 ± 1.9 pS ($18 \pm 3\%$). *b*, Single-channel currents in a patch from a post-migratory cell in the IGL (-70 mV; 10 μ M NMDA, 3 μ M glycine); the mean conductances are 50 pS (85%) and 40 pS (15%). For 21 patches from internal cells the mean single-channel conductances were 49.9 ± 0.5 pS ($83 \pm 1\%$) and 40.0 ± 0.4 pS ($17 \pm 1\%$). *c*, 'Low-conductance' single-channel currents recorded from a granule cell in the IGL (-70 mV; 50 μ M NMDA, 3 μ M glycine); this patch contained predominantly openings of 18 and 30 pS, and only these low-conductance events are shown. Similar low-conductance channels were seen in a total of 18 patches. Beyond P16 a

marked decrease in recording stability was seen for both patch and whole-cell recordings. In only four patches were the recordings sufficiently long-lived and stable to allow a full quantitative analysis. In these patches the mean conductance values were 33.0 ± 2.3 pS ($86 \pm 3\%$) and 20.3 ± 2.1 pS ($14 \pm 3\%$). For illustration, records were filtered at 1 kHz (*a* and *b*) or 2 kHz (*c*).

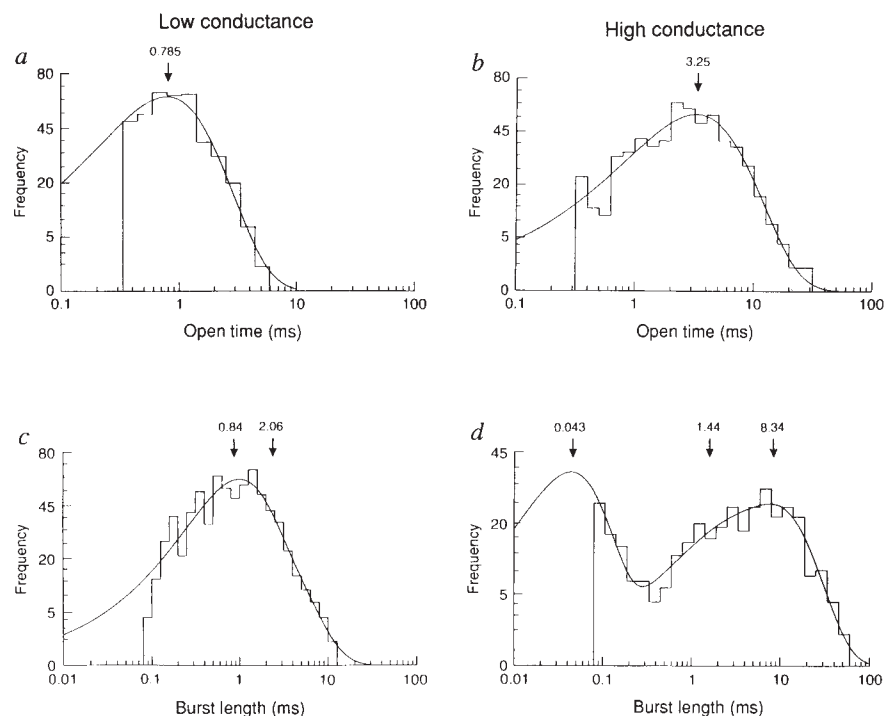
METHODS. Single-channel recordings were made from membrane patches in the outside-out configuration using either an Axopatch-1A (Axon Instruments) or L/M-EPC7 (List) amplifier. Currents were recorded on FM tape (Racal Store 4; bandwidth DC to 5.0 kHz) or on digital audio tape (BioLogic DTR-1204; DC to 20 kHz). Records were filtered at 2 kHz (8-pole Bessel filter), digitized at 20 kHz (PDP-11/73 computer with CED 502 interface) and individual currents fitted by the response function of the recording system to a step input²⁹. Mean current levels were determined from maximum likelihood fits of the sum of two to four gaussian distributions to the cursor-fitted amplitudes²⁹. Only openings longer than two filter risetimes (332 μ s) were included. The slope conductance of the main open state in internal granule cells (50.3 ± 2.6 pS, $n=3$) was constant over the voltage range investigated. The interpolated reversal potential was close to 0 mV, and this value was used to estimate chord conductances.

cells of older animals (Fig. 2*c*). These channels had mean conductance states of 20 and 33 pS (see Fig. 2 legend). Such 'low-conductance' NMDA channel openings were seen only after P13. In P19–P23 animals, when granule cell migration is nearing completion¹⁰, 65% of patches displayed low-conductance openings, either together with 50/40-pS events ($n=13$) or in isolation ($n=4$); in nine patches we detected only 50/40-pS events. As expected for NMDA channels, these low-conductance openings were reversibly blocked by AP5 (50 μ M) but not by the non-NMDA receptor antagonist CNQX (6-cyano-7-nitroquinoxaline-2,3-dione; 5–10 μ M). Figure 2*c* shows single-channel records representative of this type of response; some 50-pS events were present in this patch (see histogram), but about 80% of the channel openings were to levels of 30 and 18 pS. We conclude that these low-conductance events arose from a distinct form of NMDA channel, rather than from two separate low-conductance channels or from additional multiple conductance

levels of the 50 pS channel for the following reasons. First, direct transitions were observed between the 33- and 20-pS levels (see Fig. 2*c*), and between the 50- and 40-pS levels (Fig. 2*a, b*), but not between these high- and low-conductance events. Second, some patches from older animals exhibited low-conductance NMDA channels in the absence of 50/40-pS openings. This indicates that the high- and low-conductance channels can occur independently, and suggests that an increased expression of low-conductance channels may be accompanied by a decrease in that of 50/40-pS channels. Third, low-conductance channels showed markedly different kinetic properties from those of the 50/40-pS channels. This is evident in Fig. 2, where openings of the low-conductance channels are clearly briefer than those of the 50-pS events. The mean apparent open time of the 33-pS conductance state was 0.85 ± 0.15 ms ($n=4$), compared with 3.4 ± 0.4 ms ($n=6$) for the 50-pS channels (Fig. 3*a, b*). Similarly, the durations of the elementary events resulting from single

FIG. 3 Kinetic properties of high- and low-conductance NMDA channels in cells of the inner granular layer. Representative distributions of apparent open times and burst lengths for the main conductance states of '33-pS' channel openings (a, c) and '50-pS' channel openings (b, d), recorded in response to NMDA. Data are from four different internal granule cells. Open-time distributions for the low- and high-conductance channels are fitted with single exponentials, yielding mean open times of 0.79 and 3.25 ms, respectively. For low-conductance channels, distributions of burst lengths were best fitted with the sum of two exponentials, with mean time constants of 0.6 and 2.1 ms ($n=2$); in a third patch the distribution was best fitted with a single exponential (time constant, 1.2 ms). For high-conductance channels the distributions were best fitted with the sum of three exponentials, with time constants of $63 \pm 9 \mu\text{s}$ ($36 \pm 6\%$), $3.5 \pm 0.7 \text{ ms}$ ($32 \pm 6\%$) and $11.3 \pm 1.3 \text{ ms}$ ($32 \pm 2\%$) ($n=6$).

METHODS. Single-channel currents were digitized as described for Fig. 2. Duration of open times was determined by the method of time-course fitting²⁹. Histograms were binned logarithmically and a square root transformation of the ordinate (events per bin) was used. Exponential functions were fitted by the maximum likelihood method²⁹. Plotted in this way, peaks correspond to the time constant of the exponentials³⁰. Before fitting exponential functions to the open-time distributions a resolution of two filter risetimes ($332 \mu\text{s}$) was imposed so as to restrict the analysis to fully resolved openings to the different current amplitudes. Bursts of openings were defined as openings separated by closed times shorter than a critical length t_c . This was calculated so as to make the percentage of long closed times that were misclassified as within bursts equal to the percentage of short closed times that were misclassified as being between bursts. For each patch, t_c was



calculated from the best-fit parameters of the closed-time distribution (five components), treating components 2 and 3 as being 'within-burst' closed times. The identification of closed times between low-conductance openings was hampered by the presence of 50/40-pS events, which prevented the unambiguous assignment of brief openings to low-conductance channels. This may affect the closed-time distributions and hence the determination of t_c .

receptor activations (the burst lengths) were considerably shorter for the low-conductance NMDA channels (Fig. 3c, d).

Changes in the open times and burst lengths of NMDA channels may underlie the developmental alterations in macroscopic channel gating properties and the time course of NMDA synaptic currents seen in other brain regions^{23,24}. It is possible that these alterations result from changes in the subunit composition of NMDA receptors, as reported for acetylcholine²⁵ and glycine²⁶ receptors at developing synapses. Our observation that both high- and low-conductance NMDA channels are present in the extrasynaptic membrane suggests that both may also be present in the postsynaptic membrane at this stage. However, the low conductance and fast kinetics of the 33/20-pS channels would make their detection difficult during the excitatory postsynaptic current, unlike the high-conductance channels which can be resolved directly during the NMDA component excitatory postsynaptic currents in immature granule cells⁴. A main single-channel conductance state of 50 pS, as described here for pre-migratory and early post-migratory granule cells, is characteristic of many native NMDA-receptor channels¹²⁻¹⁶, and is observed¹⁷ when recombinant NMDA receptors (NR) are formed by co-expression of NR1 and NR2A or NR2B subunits^{7,8}. In contrast, the conductance levels and apparent open times of the 33/20-pS channels in older post-migratory granule cells are strikingly similar to those reported for the recombinant NMDA channels of the NR1-NR2C type¹⁷. This, together with evidence that NR2C subunit mRNA appears postnatally and predominantly in the cerebellum^{6,8}, suggests that the 33/20-pS channel we have identified contains the NR2C subunit. A reduced block by Mg^{2+} has been described for recombinant receptors containing this subunit^{7,8}. The apparently unaltered Mg^{2+} sensitivity of the whole-cell NMDA current that we see in older (P22) granule cells would be consistent with the fact

that cells at this stage still possess 50/40-pS channels, which because of their kinetic properties and large conductance may be expected to carry a disproportionate amount of the current. The restricted expression of the NR2C subunit⁶⁻⁸ may explain why comparable low-conductance NMDA channels have not been described in other brain regions. It is possible that 50/40-pS NMDA channels have a developmental role in young granule cells, being replaced by 33/20-pS channels at synapses in mature cells. Any change in the kinetic properties or Mg^{2+} sensitivity^{7,8} of the synaptic NMDA channels is likely to affect the time course of the NMDA component of the synaptic current and the membrane potential range over which it could normally operate.

Note added in proof: An increase in tonic NMDA-receptor activation during granule cell migration has recently been described³¹. □

Received 23 September; accepted 22 December 1993.

- Komuro, H. & Rakic, P. *Science* **260**, 95-97 (1993).
- Rabacchi, S., Bailly, Y., Delhaye-Bouchaud, N. & Mariani, J. *Science* **256**, 1823-1825 (1992).
- Garthwaite, J. & Brodbelt, A. R. *Neuroscience* **29**, 401-412 (1989).
- Silver, R. A., Traynelis, S. F. & Cull-Candy, S. G. *Nature* **355**, 163-166 (1992).
- D'Angelo, E., Rossi, P. & Taglietti, V. *Neuroscience* **53**, 121-130 (1993).
- Watanabe, M., Inoue, Y., Sakimura, K. & Mishina, M. *Neuroreport* **3**, 1138-1140 (1992).
- Kutsuwada, T. et al. *Nature* **358**, 36-41 (1992).
- Monyer, H. et al. *Science* **256**, 1217-1221 (1992).
- Ishii, T. et al. *J. Biol. Chem.* **268**, 2836-2843 (1993).
- Altman, J. J. *comp. Neurol.* **145**, 465-514 (1972).
- Hamori, J. & Somogyi, J. J. *comp. Neurol.* **220**, 365-377 (1983).
- Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
- Cull-Candy, S. G. & Usowicz, M. M. *Nature* **325**, 525-528 (1987).
- Jahr, C. E. & Stevens, C. F. *Nature* **325**, 522-525 (1987).
- Howe, J. R., Cull-Candy, S. G. & Colquhoun, D. *J. Physiol., Lond.* **432**, 143-202 (1991).
- Gibb, A. J. & Colquhoun, D. *J. Physiol., Lond.* **456**, 143-179 (1992).
- Stern, P., Béhé, P., Schoepfer, R. & Colquhoun, D. *Proc. R. Soc. Lond.* **B250**, 271-277 (1992).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85-100 (1981).

19. Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261–263 (1984).
20. Garthwaite, G., Yamini, B. Jr & Garthwaite, J. *Devl Brain Res.* **36**, 288–292 (1987).
21. Blanton, M. G. & Kriegstein, A. R. *J. Neurophysiol.* **67**, 1185–1200 (1992).
22. Sah, P., Nicoli, R. A. & Hestrin, S. *Science* **246**, 815–818 (1989).
23. Carmignoto, G. & Vicini, S. *Science* **258**, 1007–1011 (1992).
24. Hestrin, S. *Nature* **357**, 686–689 (1992).
25. Mishina, M. *et al.* *Nature* **321**, 406–411 (1986).
26. Takahashi, T., Momiyama, A., Hirai, K., Hishinuma, F. & Akagi, H. *Neuron* **9**, 1155–1161 (1992).
27. Johnson, J. W. & Ascher, P. *Nature* **325**, 529–531 (1987).
28. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. *Pflügers Arch.* **414**, 600–612 (1989).
29. Colquhoun, D. & Sigworth, F. J. in *Single-Channel Recording* (eds Sakmann, B. & Neher, E.) 191–263 (Plenum, London, 1983).
30. Sigworth, F. J. & Sine, S. M. *Biophys. J.* **52**, 1047–1054 (1987).
31. Rossi, D. J. & Slater, N. T. *Neuropharmacology* **32**, 1239–1248 (1993).

ACKNOWLEDGEMENTS. This work was supported by the Wellcome Trust, the Japanese Ministry of Education, Science and Culture and the Howard Hughes Medical Institute (International Scholars Award to S.G.C.C.). We thank D. Attwell, B. Clark, D. Colquhoun, B. Edmonds and A. Silver for discussion and comments on the manuscript.

Impaired immune and acute-phase responses in interleukin-6-deficient mice

Manfred Kopf^{*†}, Heinz Baumann[‡], Giulia Freer[§],
Marina Freudenberg^{*}, Marinus Lamers^{*},
Tadamitsu Kishimoto^{||}, Rolf Zinkernagel[§],
Horst Bluethmann[¶] & Georges Köhler^{*†}

^{*} Max-Planck-Institut für Immunbiologie, Stübweg 51,
79108 Freiburg, Germany

[‡] Department of Molecular and Cellular Biology,
Roswell Park Cancer Institute, Buffalo, New York 14263, USA

[§] Institut für Experimentelle Immunologie, Universität Zürich,
Switzerland

[¶] Department of Biology, Hoffmann-La Roche AG, CH-4002 Basel,
Switzerland

^{||} Department of Medicine III, Osaka University, Medical School,
Osaka 565, Japan

INTERLEUKIN-6 (IL-6) is a multifunctional cytokine that regulates various aspects of the immune response, acute-phase reaction and haematopoiesis (for reviews see refs 1, 2). *In vitro*, leukaemia inhibitory factor, oncostatin M, ciliary neurotrophic factor and interleukin-11 display overlapping activities with IL-6. This functional redundancy may be explained by the interactions of specific binding receptors with a common signal-transducing receptor (gp130) (for reviews see refs 3, 4). To elucidate the unique function of IL-6 *in vivo*, we have disrupted the IL-6 gene by homologous recombination. IL-6-deficient mice develop normally. They fail to control efficiently vaccinia virus and infection with *Listeria monocytogenes*, a facultative intracellular bacterium. The T-cell-dependent antibody response against vesicular stomatitis virus is impaired. Further, the inflammatory acute-phase response after tissue damage or infection is severely compromised, whereas it is only moderately affected after challenge with lipopolysaccharide. We conclude that IL-6 production induced by injury or infection is an important *in vivo* SOS signal which coordinates activities of liver cells, macrophages and lymphocytes.

The IL-6 gene was disrupted in the second exon (first coding exon) by insertion of a *neo^r* cassette (Fig. 1a). F₁ mice (C57Bl/6 × 129 Sv) heterozygous for the mutation (IL-6^{+/-}) were interbred to obtain mice homozygous for the mutation. Loss of the wild-type alleles and IL-6 messenger RNA was confirmed by Southern blot analysis and by reverse transcriptase polymerase chain reaction (RT-PCR) from lipopolysaccharide (LPS)-stimulated macrophages, respectively (not shown). Consequently, neither a bioassay nor an enzyme-linked immunosorbent assay (ELISA) with serum of LPS-treated mice indicated an IL-6 protein (Fig. 1b).

Although IL-6 was reported to be expressed as early as the eight-cell stage embryo⁵, the F₁ was generated in the expected mendelian pattern. Interbreeding of IL-6-deficient mice resulted in the normal number of pups, thus ruling out a crucial effect of IL-6 for embryogenesis. The development of the lymphoid compartment was investigated in 6–12-week-old mice by cytofluorimetric analysis. Numbers of thymocytes and peripheral T cells in IL-6^{-/-} mice were consistently reduced by 20–40% as compared to controls, suggesting that IL-6 is involved in T-cell proliferation^{6,7}. But thymocytes and peripheral T cells had normal expression patterns of T-cell receptor α , β , γ and δ chains, CD4, CD8, Pgp1 and HSA. B cells in the bone marrow and spleen of IL-6^{-/-} mice expressed B220, IgM, IgD and CD23 within normal ranges (not shown). We conclude that lymphoid development is not seriously affected in the absence of IL-6.

We next assessed T- and B-cell function in IL-6^{-/-} mice by infection with vesicular stomatitis virus (VSV) and vaccinia virus (VV). VSV infection is controlled by virus-neutralizing T-cell help-independent IgM and T-cell help-dependent IgG antibodies⁸. Virus neutralizing IgG was 5–10-fold reduced in IL-6^{-/-} mice during the course of VSV infection (Fig. 2a), whereas early (day 4) IgM titres are comparable to controls (not shown). These results are consistent with a reduced specific IgG response (affecting all IgG isotypes) against a soluble protein antigen (not shown). IL-6 was originally identified as a lymphokine inducing final maturation of B cells into antibody-secreting cells^{1,2}. IL-6-overexpressing transgenic mice develop plasmacytosis and a massive increase in polyclonal IgG1 (ref. 9). Determination of plasma cell number and natural serum antibody levels in 8–12-week-old mice, however, did not reveal differences between IL-6^{-/-} and control mice (not shown). During an antibody response dependent on T-cell help, when B cells enter the germinal centre, IL-6 is secreted by a germinal centre cell and appears to improve the expansion of plasma blasts either directly or by T-helper cells (M. Kosco-Vilbois, personal communication). Furthermore, IL-6^{-/-} mice were defective in mucosal IgA responses, which can be restored by local application of IL-6 expressed by a recombinant vaccinia virus (A. Ramsay *et al.*, manuscript in preparation). Thus, although local IL-6 production can increase B-plasma cell formation there are additional ways to stimulate B cells because IgM responses were not affected and IgG responses were only reduced but not absent.

The generation and activity of cytotoxic T cells (CTL) against vaccinia virus (WR strain) was 3–10-fold reduced in IL-6^{-/-} mice (Fig. 2b). By contrast, CTL function against lymphocytic choriomeningitis virus (LCMV) and an attenuated vaccinia virus recombinant strain was comparable to controls (not shown). Differences in CD4⁺ T-cell-mediated help for CD8⁺ T-cell responses against vaccinia virus as compared to LCMV may account for this^{10,11}. The defence against vaccinia virus (WR) in IL-6^{-/-} mice was compromised. Virus titres were 10–1,000-fold augmented in ovaries and lungs of IL-6^{-/-} mice (Fig. 2c), which were severely cachectic when killed for virus titre determination. Because β_2 -microglobulin-deficient mice lacking CD8⁺ T cells were able to control vaccinia virus¹², and because interferon- γ (IFN- γ) and tumour necrosis factor- α (TNF- α) are crucial factors for virus clearance^{13,14}, our result implies that these or other factors with a direct or indirect antiviral activity may be influenced by IL-6.

Infection with *L. monocytogenes* was studied to evaluate T-cell and macrophage interaction¹⁵. Infection of IL-6^{-/-} mice resulted in a 100–1,000-fold higher listerial titre in spleen and liver as compared to controls (Fig. 3a). This may be due to reduced bactericidal activity of the macrophage, which is known to be mediated by release of toxic nitrogen intermediates (for review see ref. 16). Measurement of nitrite release in thioglycollate-elicited peritoneal macrophages after *in vitro* stimulation with LPS and IFN- γ , however, did not show any difference between IL-6^{-/-} and control mice (Fig. 3b). Thus, macrophages can be stimulated by LPS-induced TNF- α and IFN- γ in the absence of

[†] To whom correspondence should be addressed.

FIG. 1 Generation of an IL-6-deficient mouse. *a*, Targeting strategy. The IL-6 target vector with a short arm and long arm homology of 1.4 kilobases (kb) and 8.5 kb, respectively, contains the *neo* cassette cloned in exon 2 of the IL-6 gene together with the HSVtk cassette flanking the 3' homology. Genomic DNA of homologous recombination PCR-positive D3 embryonic stem cell clones were digested with both *Eco*RI and *Stu*I. Southern hybridization with probes A and *neo* was consistent with disruption of the IL-6 wild-type allele (not shown). *b*, IL-6 detection. IL-6-deficient and control mice were injected intraperitoneally (●) or intravenously (■) with 25 µg LPS (Sigma). Blood was taken after 4 h. Left, An IL-6 bioassay was performed by incubating the IL-6-dependent B-cell hybridoma 7TD1 (7×10^3 cells per microwell) with serial serum dilutions of both groups of mice. Cell growth was evaluated by colorimetric measurement of hexosaminidase levels²⁵. Given is the serum titre ($\log_{10}$)⁻¹ at half maximal absorbance of 1.3. Right, In parallel, serum of mice was assayed for IL-6 by ELISA and standardized with mIL-6 (EM-IL6 kit, Endogen).

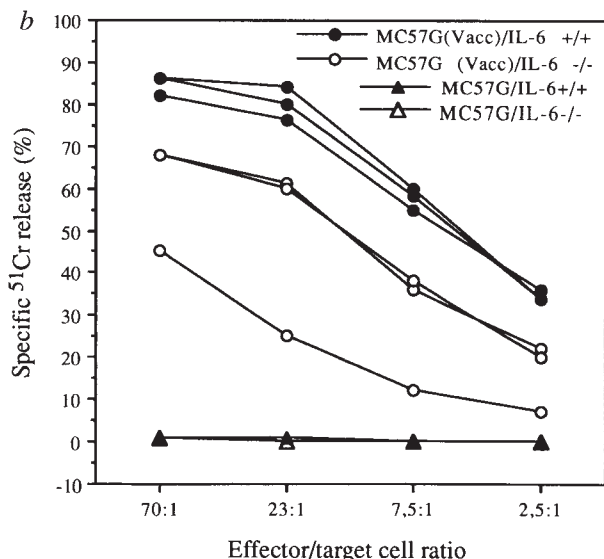
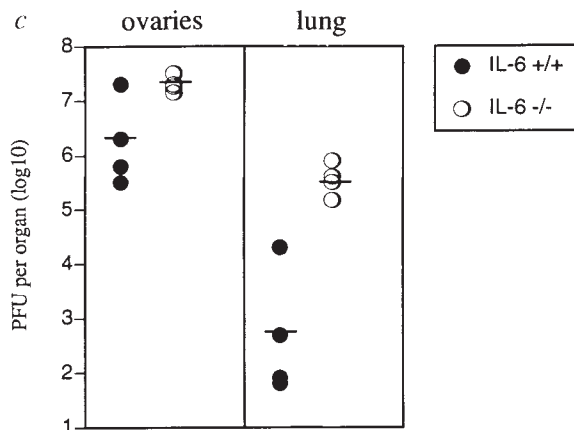
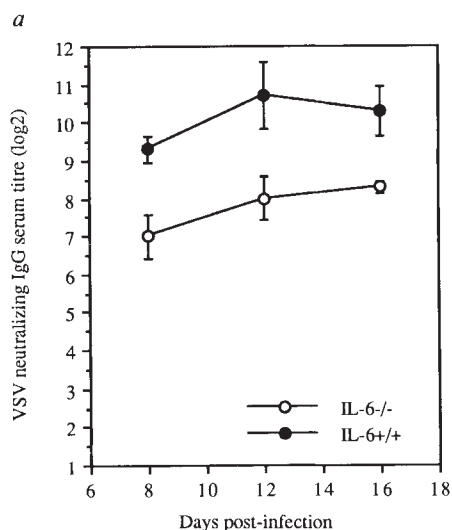
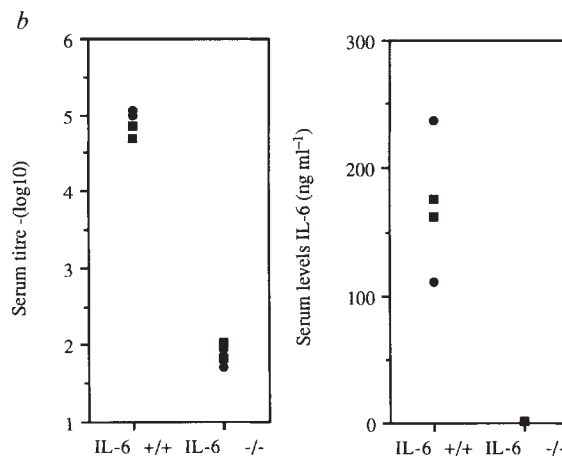
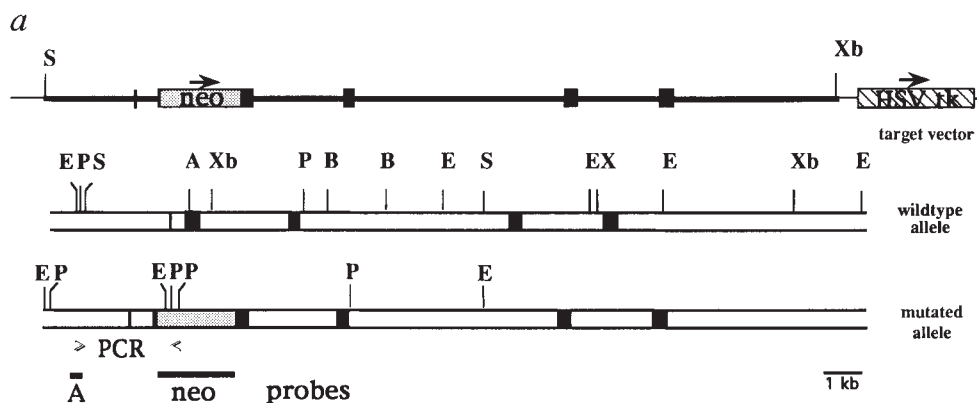


FIG. 2 Defence against viral infections. *a*, Anti-vesicular stomatitis virus (VSV) IgG response. Mice ($N = 4$ per group) were infected intravenously with 2×10^6 plaque-forming units (PFU) VSV (Indiana strain). Blood samples were collected at day 4, 8, 12 and 16 after infection. Neutralizing anti-VSV IgM (not shown) and IgG antibody titres were defined as the highest serum dilution that reduced the number of plaques to half of the control. The titre of IgG was determined by twofold dilution steps of sera starting with 1:40 and reduction of IgM with 0.1 M 2-mercaptoethanol. Values represent mean titres of three mice ($\pm$ s.e.). One mouse of each group died during infection. *b*, Cytotoxic T-cell responses. Mice were infected with 10^6 (PFU) of vaccinia virus (WR strain). T cells were isolated from the spleens at day 6 and assayed for specific lysis of MC57G (H-2^b) target cells infected (circles) or uninfected (triangles) with vaccinia virus. The percentage specific release of ⁵¹Cr was calculated as (experimental release - spontaneous release) $\times$ 100/(total release - spontaneous release). IL-6^{+/+} closed symbols; IL-6^{-/-} open symbols. *c*, Vaccinia virus titre. The number of vaccinia virus recovered from ovaries and lungs from mice killed for CTL assay (*b*) at day 6 post-infection²⁶.

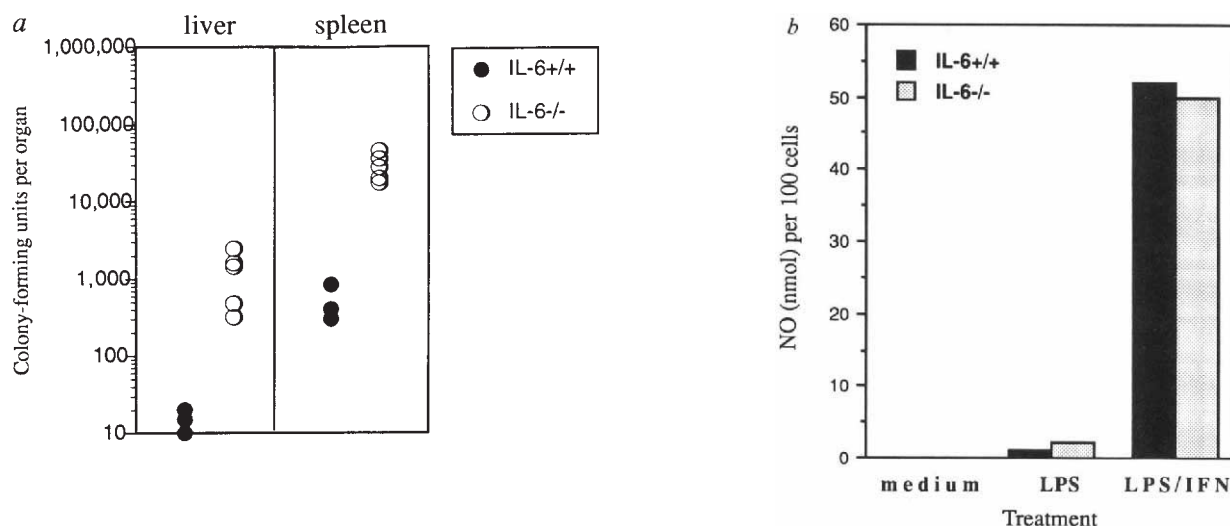


FIG. 3 *a*, Infection with *L. monocytogenes*. Mice were infected intravenously with 3.0×10^3 colony forming units (CFU). Numbers of viable bacteria recovered from the spleen and liver were determined at day 5 post-infection by plating serial 10-fold dilutions of organ homogenates in PBS on Trypticase-soy agar²⁷. The values of individual mice ($N=5$ per group) is shown. A second experiment gave similar results (not shown). *b*, Nitric oxide production. Thioglycollate-elicited peritoneal macrophages were left unstimulated or stimulated with LPS (100 ng ml^{-1}) in the absence or presence of IFN- γ (100 ng ml^{-1}). After 48 h, supernatants were determined for nitric oxide content by measurement of nitrite generated from nitric oxide using the Griess reagent²⁸. Values represent duplicate measurements. Closed bars: IL-6 $^{+/+}$ mice; open bars: IL-6 $^{-/-}$ mice.

METHODS. Mice ($N=3$ per group) were injected intraperitoneally with 2 ml 4% thioglycollate broth. At day 4 peritoneal cavity cells were collected, pooled per treatment group, and incubated for 1 h at 37°C at a density of 10^6 cells per ml Iscove medium containing Glutamin (2 mM) and 2-mercaptoethanol (10^{-4} M). Nonadherent cells were removed. Adherent cells were stimulated as indicated. After 48 h the supernatants (100 μl) were mixed with Griess reagent (100 μl) and the absorbance measured at 565 nm. Sodium nitrite was used as standard. Acute-phase response (see Table 1): mice ($N=3$ per group) were treated either by (1) intravenous injection of *L. monocytogenes* (3×10^3 CFU); or (2) subcutaneous injection of 50 μl sterile turpentine (Oleum Terebinthinae; Mainland Pharmazeutische Fabrik, Frankfurt); or (3) intravenous injection of 1 μg LPS (Sae); or 10 μg hIL-6 (kindly provided by L. Aarden) together with 2–5 μg dexamethasone (Sigma). Mice were killed after indicated time points. Total liver RNAs were prepared and subjected to northern blot analysis using as probes complementary DNA for mouse Haptoglobin (HP), α 1-acid glycoprotein (AGP) and serum amyloid A (SAA). α 1-Antitrypsin (AT) was used as an internal marker, a plasma protein not affected by inflammation in the mouse²⁹. The hybridization signals were quantified by phosphorimaging. The normalized hybridization signals for AGP and HP were expressed relative to the values of untreated IL-6 $^{+/+}$ mice ($N=2$). Because normal mice have extremely low levels of SAA mRNA, the level of IL-6 $^{+/+}$ mice treated with turpentine was arbitrarily defined as 100. The amounts of AGP and HP were determined by rocket immunoelectrophoresis of equal aliquots of serum on double-section agarose gels containing antibodies³⁰.

IL-6, *in vitro*. Whether such a stimulation occurs in the *Listeria*-infected macrophages *in vivo* remains to be seen^{17,18}. Macrophages become listericidal only after T-cell activation, which could be impaired in IL-6-deficient mice¹⁹. An inability to clear intracellular bacterial pathogens was also observed in mice deficient for either IFN- γ ²⁰, IFN- γ -receptor¹³, or TNF-receptor^{21,22} after gene targeting, or for IL-1-receptor after antibody treatment²³. IL-6 may, therefore, be involved in optimally stimulating the activity of these factors in an unknown way.

The acute-phase response is mediated by several factors including IL-6, IL-1, TNF and leukaemia inhibitory factor

(LIF)²⁴. To elucidate the unique role of IL-6 in the liver acute-phase response, IL-6 $^{-/-}$ and control mice were challenged either with turpentine, LPS, or bacterial infection. In IL-6 $^{+/+}$ mice, both a local sterile tissue damage by subcutaneous injection of turpentine, and infection with Gram-positive *L. monocytogenes* induced a remarkable increase in liver haptoglobin (HP), α 1 acid glycoprotein (AGP) and serum amyloid A (SAA) RNA (Table 1). No such increase was observed in IL-6 $^{-/-}$ mice. The most striking difference (100-fold) between mutant and control mice was measured for SAA mRNA. By contrast, the acute-phase response of IL-6 $^{-/-}$ mice treated with LPS was only slightly reduced as compared to controls (Table 1). Serum levels of HP and AGP (SAA not determined) were in line with the mRNA levels. Injection of IL-6 induced an equivalent response in both groups of mice, demonstrating that liver cells of IL-6 $^{-/-}$ mice are still responsive to IL-6. Thus, IL-6 is an important mediator of an acute-phase response after tissue damage or infection with gram-positive intracellular bacteria but not Gram-negative bacteria. Bacterial LPS has systemic rather than local effects and triggers inflammatory mediators from a variety of cell types and may therefore be less dependent on IL-6.

The data show that optimal responses to trauma and infection can only be mediated in the presence of IL-6, despite an apparent functional redundancy of factors with IL-6-like activities. □

TABLE 1 Acute-phase protein response

Treatment	Haptoglobin		α 1-AGP		SAA	
	IL-6 $^{+/+}$	IL-6 $^{-/-}$	IL-6 $^{+/+}$	IL-6 $^{-/-}$	IL-6 $^{+/+}$	IL-6 $^{-/-}$
Control						
mRNA	1.0	1.2	1.0	0.7	<0.2	<0.2
Serum (mg ml $^{-1}$)	<0.1	<0.1	0.4	0.3	ND	ND
Turpentine 24 h						
mRNA	23.7	4.5	48.4	3.9	100	1
Serum (mg ml $^{-1}$)	4.0	0.9	1.2	0.3	ND	ND
<i>Listeria</i> 48 h						
mRNA	13.9	3.9	32.9	5.2	40	1
Serum (mg ml $^{-1}$)	2.6	1.2	1.0	0.4	ND	ND
LPS 26 h						
mRNA	16.0	7.6	27.1	10.8	55	25
Serum (mg ml $^{-1}$)	1.7	0.7	1.2	0.7	ND	ND
IL-6 24 h						
mRNA	4.1	8.3	1.8	7.7	12	14
Serum (mg ml $^{-1}$)	0.5	0.9	0.6	0.6	ND	ND

Received 20 September 1993; accepted 8 February 1994.

1. Van-Snick, J. A. *Rev. Immun.* **8**, 253–278 (1990).

2. Kishimoto, T. *Blood* **74**, 1–10 (1989).

3. Kishimoto, T., Akira, S. & Taga, T. *Science* **258**, 593–597 (1992).

4. Stahl, N. & Yancopoulos, G. D. *Cell* **74**, 587–590 (1993).
5. Rothstein, J. L. et al. *Genes Dev.* **6**, 1190–1201 (1992).
6. Lotz, M. et al. *J. exp. Med.* **167**, 1253–1258 (1988).
7. Uytendhoeve, C., Coulie, P. G. & Van, S. J. *J. exp. Med.* **167**, 1417–1427 (1988).
8. Gupta, S. L., Hengartner, H. & Zinkernagel, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2604–2608 (1986).
9. Suematsu, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7547–7551 (1989).
10. Leist, T. P., Kohler, M. & Zinkernagel, R. M. *Scand. J. Immun.* **30**, 679–686 (1989).
11. Doherty, P. C., Allan, W. & Eichelberger, M. A. *Rev. Immun.* **10**, 123–151 (1992).
12. Spriggs, M. K. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6070–6074 (1992).
13. Huang, S. et al. *Science* **259**, 1742–1745 (1993).
14. Ramsay, A. J., Ruby, J. & Ramshaw, I. A. *Immun. Today* **14**, 155–157 (1993).
15. Kaufmann, S. A. *Rev. Immun.* **11**, 129–163 (1993).
16. Liew, F. Y. & Cox, F. E. G. *Immunoparasit. Today* A17–A21 (1991).
17. Beckerman, K. P. et al. *J. Immun.* **150**, 888–895 (1993).
18. Gregory, S. H., Wing, E. J., Hoffman, R. A. & Simmons, R. L. *J. Immun.* **150**, 2901–2909 (1993).
19. North, R. J. *J. Immun.* **121**, 806–810 (1978).
20. Dalton, D. K. et al. *Science* **259**, 1739–1742 (1993).
21. Pfeffer, K. et al. *Cell* **73**, 457–467 (1993).
22. Rothe, J. et al. *Nature* **364**, 798–802 (1993).
23. Havel, E. A., Moldawer, L. L., Helfgott, D., Kilian, P. L. & Sehgal, P. B. *J. Immun.* **148**, 1486–1492 (1992).
24. Koj, A. in *Injury and Infection* (eds Gordon, A. H. & Koj, A.) 145–160 (Elsevier, Amsterdam, 1985).
25. Van Snick, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9679–9683 (1986).
26. Binder, D. & Kundig, T. M. *J. Immun.* **146**, 4301–4307 (1991).
27. Blanden, R. V. & Langmann, R. E. *Scand. J. Immun.* **1**, 379–391 (1972).
28. Ding, A. H., Nathan, C. F. & Stuehr, J. *Immun.* **141**, 2407–2412 (1988).
29. Baumann, H., Latimer, J. J. & Glibetic, M. D. *Arch. Biochem. Biophys.* **246**, 488–493 (1986).
30. Baumann, H. *Meth. Enzym.* **163**, 566–594 (1988).

ACKNOWLEDGEMENTS. We thank Y. Lang for blastocyst injection, K.-H. Widmann for embryo transfer and extensive work in the animal facilities, Y. Wang for mRNA analysis, and P. Koebel and N. Schultze for technical help.

Biological activity of soluble wingless protein in cultured *Drosophila* imaginal disc cells

Frank van Leeuwen, Cynthia Harryman Samos & Roel Nusse*

Howard Hughes Medical Institute and Department of Developmental Biology, Beckman Center, Stanford University, School of Medicine, Stanford, California 94305, USA

THE phenotypes caused by mutations in *Wnt* genes suggest that their gene products are involved in cell-to-cell communication^{1–6}. *Wnt* genes indeed encode secreted molecules^{7–10}, but soluble active *Wnt* protein has not been found. We have developed a novel cell culture assay for the *Drosophila Wnt* gene *wingless*^{11,12}, using a *Drosophila* imaginal disc cell line (cl-8; ref. 13), and measured effects on the adherens junction protein armadillo^{14,15}, a known genetic target of *wingless*¹⁶. Transfection of a temperature-sensitive *wingless* complementary DNA into cl-8 cells increases the levels of the armadillo protein. The *wingless* protein does not affect the rate of synthesis of armadillo, but leads to increased stability of an otherwise rapidly decaying armadillo protein. The *wingless* protein in the extracellular matrix and soluble medium from donor cells also increases the levels of armadillo protein. The protein in the medium acts fast and is inhibited by an antibody to *wingless* protein, demonstrating that *Wnt* products can act as soluble extracellular signalling molecules.

We transfected cl-8 cells with expression constructs containing a temperature-sensitive *wingless* protein (Wg) mutant, *wg*^{IL114}, and constructs with two non-conditional lethal alleles, *wg*^{IN67} and *wg*^{CX2} (ref. 17). *In vivo*, the *wg*^{IL114} allele behaves as wild type at 16–18 °C and as a non-functional gene at 25–29 °C (refs 17, 18). We found a roughly 10-fold increase in armadillo (Arm) protein levels at 16 °C compared to 29 °C in the cells expressing *wg*^{IL114} (Fig. 1a). The cells expressing the non-conditional mutant alleles contained levels of Arm protein similar to untransfected cells at both temperatures.

The temperature-dependent effect of Wg allowed us to study

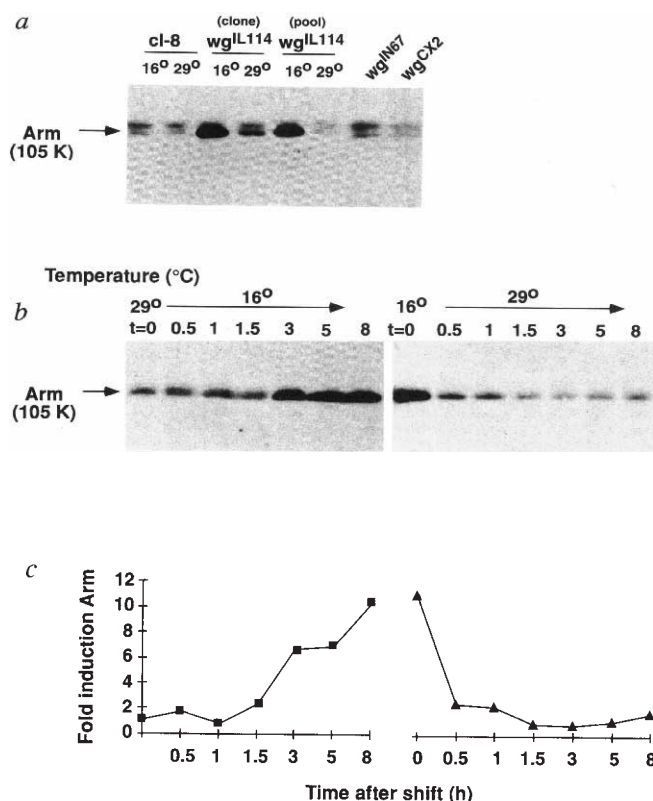


FIG. 1 a, Arm protein levels in the various transfected cl-8 cells detected by western blotting. Arm accumulates in cells expressing the temperature-sensitive *wg*^{IL114} protein at the permissive temperature of 16 °C but not at 29 °C in both a clonal cell line and in pools of transfected cells when compared to untransfected cl-8 cells. Pools expressing *wg*^{IN67} or *wg*^{CX2} have normal Arm levels. The Arm protein migrates as two bands around 105K, reflecting a difference in post-translational modification (M. Peifer and E. Wieschaus, personal communication). The faster-migrating form accumulates predominantly in the Wg-expressing cells. b, Time course of Arm protein accumulation in cl-8 cells expressing the temperature-sensitive *wg*^{IL114} allele. Shifting cells from 29 to 16 °C leads to an increase in Arm levels within 3 h after temperature shift, reaching maximum levels between 5 and 8 h (left). When cells were returned from 16 to 29 °C, the amount of Arm decreased rapidly (right). c, Arm levels at different times were quantified by laser densitometry. Numbers represent average of two independent experiments.

METHODS. The *Drosophila* wing imaginal disc cell line cl-8 (ref. 13) was cultured in Shields and Sang's MM3 medium supplemented with 100 U ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin, 2% heat-inactivated FCS, 0.125 i.u. ml⁻¹ insulin (Sigma) and 2.5% of a fly extract prepared as in ref. 25. The *wg*^{IL114} and the *wg*^{IN67} alleles give rise to single amino-acid substitutions: a serine replaces a cysteine in the *wg*^{IL114} protein and a glycine replaces an aspartic acid in the *wg*^{IN67} protein; whereas in *wg*^{CX2} protein, the last 23 amino acids of the protein are deleted¹⁷. The expression vector used, pMK33, is designed to express proteins under control of the *Drosophila* metallothionein promoter and carries a hygromycin resistance gene to allow selection of stable cell lines²⁶. Similar amounts of Wg protein were detected in lysates of cells grown in the presence or absence of copper sulphate, so this induction was consequently omitted. cl-8 cells were grown to 90% confluency, washed with PBS and lysed in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1% Nonidet-P40, 5 mM EDTA) supplemented with 20 µg leupeptin, 100 µg aprotinin and 180 µg phenylmethylsulphonyl fluoride (PMSF) per ml. We used roughly equal numbers of cells per assay. The extracts were subjected to electrophoresis and western blotting. Blots were first stained in Ponceau red to evaluate equal loading of total protein and transfer, then incubated overnight in blocking buffer with monoclonal anti-Arm antibody 7A1 (ref. 27) at a 1:1,000 dilution or a rabbit polyclonal anti-Wg antibody⁷ at a 1:5,000 dilution. The blots were washed three times for 15 min each in TBST (10 mM Tris, pH 7.5, 150 mM NaCl, 0.2% Tween-20), incubated for 1 h with horseradish peroxidase-conjugated secondary antibodies (Biorad) diluted 1:20,000 in blocking buffer.

* To whom correspondence should be addressed.

the kinetics of Arm protein elevation in a convenient way. When cells were shifted from the non-permissive temperature (29 °C) to the permissive temperature of 16 °C, Arm protein levels became higher after 3 h and reached maximum levels at 5–8 h. In the converse experiment, shifting from 16 to 29 °C, Arm protein levels were lower at 0.5 h (Fig. 1*b, c*). The temperature-dependent accumulation of Arm in the cells transfected with *wg*^{IL114} shows conclusively that accumulation of Arm is caused

by functional expression of Wg, and is not due to clonal variation between transfected cells.

The rapid decrease in Arm protein levels when Wg is inactivated suggested that the Arm protein has a rapid turnover and that the overall increase could be due to stabilization of the protein. Pulse-chase metabolic labelling experiments showed that this is probably the case. In pulse (5 min)-labelled cells, Arm protein was made in similar amounts, irrespective of the tem-

FIG. 2 Wg stabilizes Arm but does not affect the rate of Arm synthesis. *a*, left, cl-8 cells and cl-8 cells expressing *wg*^{IL114} were labelled with [³⁵S]methionine for 5 min at 16 or 29 °C and then lysed. The amount of label incorporated was determined and immunoprecipitations were from a standardized amount of radioactivity. Incorporation of radioactivity into Arm appears to be independent of the presence of functional Wg. Right, pulse-chase analysis of Arm in cl-8 cells and cells transfected with the *wg*^{IL114} construct at 16 °C. Cells were pulse-labelled for 15 min. Chase times are indicated. Arm levels are stable over 2.5 h in the presence of functional Wg but not in the control cells. *b*, Kinetics of Arm turnover in cl-8 cells and *wg*^{IL114}-expressing cells at 16 and 29 °C. Intensities were quantified using laser densitometry with *t*=0 set at 100%. Arm decays rapidly in the absence of functional Wg. **METHODS.** cl-8 cells were grown to confluency in 12-well dishes and labelled at 16 °C with 0.4 mCi of [³⁵S]methionine in 0.5 ml of M3 medium lacking methionine. After 15 min the label was removed, the monolayers were rinsed with cold complete M3 medium supplemented with 25 mM unlabelled methionine (chase). After the indicated chase times at 16 or at 29 °C, the cells were briefly rinsed with cold PBS and lysed. The lysates were then incubated for 1 h with the anti-Arm monoclonal antibody 7A1 diluted 1:50. Immune complexes were precipitated with 60 µl Protein-A sepharose beads coupled to a secondary rabbit anti-mouse antibody (Pierce), washed and analysed on a 7.5% SDS polyacrylamide gel.

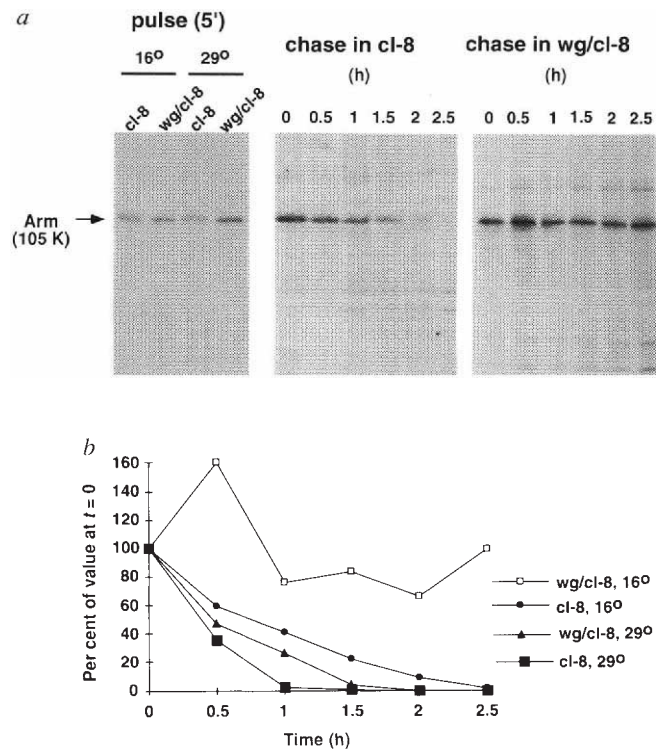
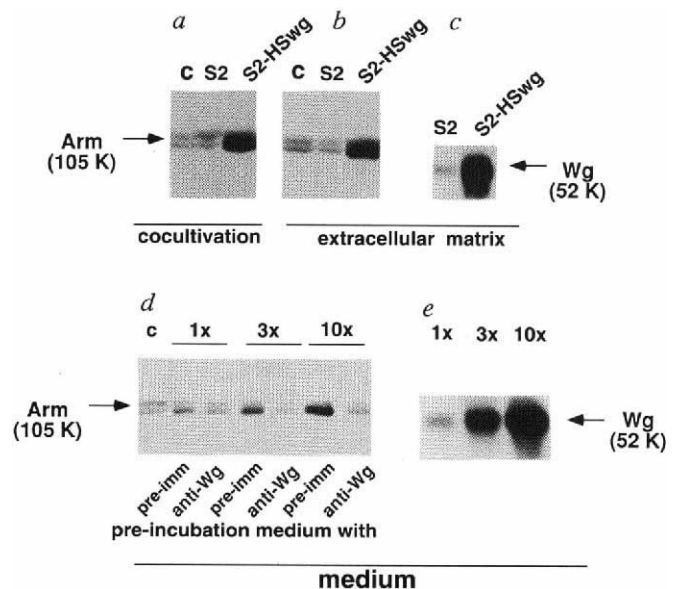


FIG. 3 cl-8 cells respond to Wg provided by S2 cells overexpressing Wg (S2-HSwg). *a*, Arm levels in cl-8 cells (*c*) compared to cl-8 cells co-cultured for 3 h in the presence of S2 cells or S2-HSwg cells. S2 cells were removed before lysis of the cl-8 cells. *b*, Arm levels in cl-8 cells (*c*) compared to cl-8 cells cultured on extracellular matrix (ECM) produced by S2 or S2-HSwg cells. *c*, In parallel, matrix material secreted by S2 or S2-HSwg cells was collected and analysed for the presence of Wg by western blot. A large amount of wingless protein is present in the matrix fraction derived from S2-HSwg cells. A faint band is also visible in the matrix fraction derived from the control S2 cells but this probably represents a protein crossreacting with the wingless antibody: control S-2 cells express no detectable Wg RNA and this band migrates slightly faster than the Wg protein¹⁷. *d*, Arm levels in cl-8 cells incubated for 3 h with medium conditioned by S2-HSwg cells. Medium fractions were either unconcentrated or concentrated 3- or 10-fold and added to cl-8 cells after pre-incubation with a rabbit polyclonal antiserum to Wg at 4% final concentration or a pre-immune serum at the same concentration. *e*, Untreated control cl-8 cells. Medium from non-transfected S2 cells showed no activity (not shown). *e*, In parallel the amount of Wg present in these medium fractions was determined by western blotting.

METHODS. cl-8 cells were grown to 80% confluency in 12-well dishes. For the co-cultivation experiments, S2-HSwg (ref. 19) or control S2 cells were heat-shocked for 30 min at 37 °C and allowed to recover for 2 h. A total of 1 ml cells ($\sim 2 \times 10^6$ cells ml⁻¹) was added to the cl-8 cells and incubated for 3 h. Then the S2 cells were removed by squirting with PBS (the S2 cells are non-adherent and can thus be separated easily) and lysates were made of the cl-8 cells. For ECM experiments, 6-well dishes were conditioned overnight by growing 2 ml S2-HSwg cells (after 30-min heatshock) or control S2 cells at a concentration of $\sim 2 \times 10^6$ cells ml⁻¹. The cells were removed the next day by washing the dish with PBS followed by scraping until no more cells were present.



cl-8 cells were seeded into the same dish and allowed to adhere for 5 h after which the cells were lysed. Conditioned medium was collected from S2-HSwg cells ($\sim 2 \times 10^7$ cells per 10 ml) for 3 h in the absence of FCS. The cells were removed by centrifugation at 2,000g. Insoluble material was then removed by ultra-centrifugation (1.5 h; 100,000g). The conditioned medium was concentrated using a Centrprep 10 column (Amicon).

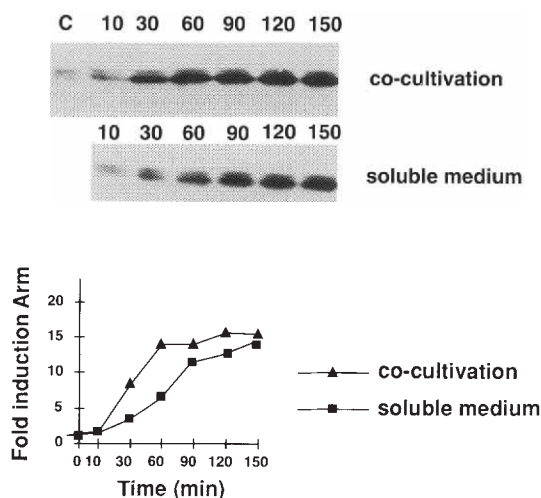


FIG. 4 Time course of Arm accumulation in response to extracellular Wg. Arm levels in cl-8 cells compared to cells co-cultured with S2-HSwg cells for time intervals ranging from 10–150 min (top). Alternatively, Arm levels were determined in cl-8 cells incubated for the same time intervals with soluble medium conditioned by S2-HSwg cells (middle). Kinetics of induction of Arm accumulation by Wg was quantified by laser densitometry (bottom).

perature at which the cells were grown or whether the cells were transfected with Wg constructs or not (Fig. 2a). To examine the turnover of the Arm protein, cells were briefly (15 min) labelled at the permissive temperature, followed by a chase/incubation in the absence of label and an excess of unlabelled methionine. Over 2.5 h the Arm protein in cells making functional Wg appeared to be stable, whereas in cells not expressing Wg, Arm protein decayed rapidly (Fig. 2a, b). Hence, the Arm protein has a short half-life (~0.5 h) and its accumulation due to Wg appears to be largely a consequence of increased stability rather than an increase in the rate of protein synthesis. As in the *Drosophila* embryo¹⁶, Arm messenger RNA levels were not affected by Wg activity (not shown).

We next asked whether Wg from sources outside the target cells could influence Arm protein levels. We first tested possible effects in a paracrine assay, co-culturing donor cells producing Wg and non-transfected cl-8 cells. As donor cells we used Wg-overexpressing Schneider 2 (S2) cells^{17,19}. The cl-8 cells that were co-cultivated with the Wg-overexpressing S2 cells showed a large increase in Arm levels (Fig. 3a).

We then examined whether the extracellular matrix (ECM) from S2 cells had Wg activity. S2 cells expressing Wg and non-transfected control S2 cells were grown overnight and removed from the dish by scraping. Figure 3c shows the presence of the

Wg protein in the material remaining on the dishes (see also ref. 17). In parallel, cl-8 cells were seeded on dishes conditioned with the ECM and allowed to adhere, leading to a large increase in Arm protein (Fig. 3b).

We then tested whether biologically active Wg could be detected in the tissue culture supernatant. Serum-free medium conditioned for 3 h by Wg overexpressing S2 cells was centrifuged at 100,000g for 1.5 h and concentrated. Analysis of a sample of the medium by blotting showed the Wg protein present in increasing concentrations in the medium (Fig. 3e). These medium fractions were then incubated with cl-8 cells for 3 h. A significant induction of Arm (Fig. 3d) corresponded to the increase in Wg concentration.

To demonstrate that the active component in the concentrated tissue culture medium from the Wg-producing S2 cells was Wg, we added to the medium a rabbit antiserum that was raised against a bacterial fusion protein containing Wg⁷ and used the pre-immune serum of the same animal as a control. After a 1-h pre-incubation, the medium was added to the cells. Figure 3d shows that the antiserum removed the Arm-inducing effect from the medium, whereas a control pre-immune serum had no effect.

Finally, we examined the kinetics of Arm accumulation, by incubating the cl-8 cells with either the S2 cells producing Wg or with concentrated soluble medium (Fig. 4). Arm levels were moderately increased after a 10 min incubation, whereas maximum levels were reached after 90 min. With the co-cultivated Wg-producing cells, 50% of the maximal induction was obtained after 30 min incubation.

Importantly, the turnover of Arm in the absence of Wg, as measured by the pulse-chase experiment shown in Fig. 2, has a similar time course: $t_{1/2} \sim 30$ min. We conclude from these similar kinetics that extracellular Wg has a rapid effect, in the order of minutes. This is because Wg does not increase the rate of synthesis of Arm, but rather inhibits its decay; the elevation in the steady-state level of Arm can therefore not be faster than the rate of degradation. This rapid effect is noteworthy in view of the genetic evidence that Arm is required for Wg signalling^{20,22}. Most if not all *in vivo* effects of Wg require the presence of functional Arm, and Arm may therefore be a component of the Wg signal transduction pathway which becomes modified and stabilized by Wg activity.

What is the nature of this modification? Wg predominantly elevates a form of Arm that migrates faster in gels, and this faster migrating form is underphosphorylated compared to the slower form (M. Peifer and E. Wieschaus, personal communication). It is possible that this differential phosphorylation is regulated by a protein kinase encoded by *zeste white-3* or *shaggy*^{23,24}. *zw-3* acts upstream of *arm* and *arm* becomes uniformly highly expressed in embryos lacking *zw-3* function^{21,22}. Thus, *zw-3* protein kinase activity may directly or indirectly phosphorylate Arm, destabilizing the protein. Wg would then inhibit *zw-3*, in turn leading to accumulation of Arm. □

Received 6 December 1993; accepted 15 February 1994.

- Nusse, R. & Varmus, H. E. *Cell* **69**, 1073–1087 (1992).
- McMahon, A. P. *Trends Genet.* **8**, 236–242 (1992).
- McMahon, A. P., Joyner, A. L., Bradley, A. & McMahon, J. A. *Cell* **69**, 581–595 (1992).
- Struhl, G. & Basler, K. *Cell* **72**, 527–540 (1993).
- DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassisi, J. & O'Farrell, P. *Nature* **332**, 604–609 (1988).
- Chu-Lagarrigues, Q. & Doe, C. *Science* **261**, 1594–1597 (1993).
- Van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. *Cell* **59**, 739–749 (1989).
- Papkoff, J. & Schryver, B. *Molec. cell. Biol.* **10**, 2723–2730 (1990).
- Bradley, R. S. & Brown, A. M. *EMBO J.* **9**, 1569–1575 (1990).
- González, F., Swales, L., Bejsovec, A., Skaer, H. & Martínez-Arias, A. *Mech. Dev.* **35**, 43–54 (1991).
- Rijsewijk, F. et al. *Cell* **50**, 649–657 (1987).
- Baker, N. E. *EMBO J.* **6**, 1765–1773 (1987).
- Peel, D. J. & Milner, M. J. *Wilhelm Roux Arch. dev. Biol.* **201**, 120–123 (1992).
- Peifer, M. & Wieschaus, E. *Cell* **63**, 1167–1178 (1990).
- McCrea, P. D., Turck, C. W. & Gumbiner, B. *Science* **254**, 1359–1361 (1991).
- Riggelman, B., Schedl, P. & Wieschaus, E. *Cell* **63**, 549–560 (1990).

- van den Heuvel, M., Harryman-Samos, C., Klingensmith, J., Perrimon, N. & Nusse, R. *EMBO J.* **12**, 5293–5302 (1993).
- Bejsovec, A. & Martínez-Arias, A. *Development* **113**, 471–485 (1991).
- Cumberledge, S. & Krasnow, M. A. *Nature* **363**, 549–552 (1993).
- Noordermeer, J., Klingensmith, J., Perrimon, N. & Nusse, R. *Nature* **367**, 80–83 (1994).
- Peifer, M., Sweeton, D., Casey, M. & Wieschaus, E. *Development* **120**, 369–370 (1994).
- Siegfried, E., Wilder, E. & Perrimon, N. *Nature* **367**, 76–80 (1994).
- Bourouis, M. et al. *EMBO J.* **9**, 2877–2884 (1990).
- Siegfried, E., Perkins, L. A., Capaci, T. M. & Perrimon, N. *Nature* **345**, 825–829 (1990).
- Currie, D. A., Milner, M. J. & Evans, C. W. *Development* **102**, 804–814 (1988).
- Koelle, M. R. et al. *Cell* **67**, 59–77 (1991).
- Peifer, M. J. *Cell Sci.* **105**, 993–1000 (1993).

ACKNOWLEDGEMENTS. We thank D. Peel and M. Milner for kindly sending us the cl-8 cells; and E. Wieschaus and M. Peifer for the anti-Armadillo monoclonal antibody 7A1 and for communication of unpublished results. We thank S. Cumberledge and M. Krasnow for the S2-HSwg cells. We thank the members of our laboratory and G. Barsh and D. Kingsley for discussions and useful comments on the manuscript. This research was supported by the Howard Hughes Medical Institute, of which R.N. is an investigator.

Evolution of RNA editing in kinetoplastid protozoa

Dmitri A. Maslov*, Herbert A. Avila*, James A. Lake*† & Larry Simpson*‡§

* Department of Biology, † Molecular Biology Institute, ‡ Howard Hughes Medical Institute, University of California, Los Angeles, 5-748 MacDonald Building, 10833 Le Conte Avenue, Los Angeles, California 90024, USA

§ To whom correspondence should be addressed

THE editing of RNA in trypanosomatid mitochondria involves the insertion and occasional deletion of uridine residues within coding regions of maxicircle messenger RNA transcripts. The extent to which the transcripts of homologous genes undergo editing differs in different species. In some, entire genes are edited (pan-editing), whereas in others, editing is limited to the 5' termini of editing domains (5' editing)^{1,2}. Here we investigate which type of editing is ancestral and which is derived, by analysing RNA editing in the different lineages, using a kinetoplastid phylogeny reconstructed from nuclear small subunit ribosomal RNA sequences. We conclude that the ancestral cryptogenes were pan-edited, and we hypothesize that the 5'-edited homologues were generated by several independent events from partially edited RNAs, in which case editing may be a more primitive mechanism than previously thought.

The order Kinetoplastida contains two major subgroups that differ in morphological characters and life cycles: the trypanosomatids and the bodonids and related cryptobiids³. RNA editing occurs in all trypanosomatid species so far examined and also in the cryptobiid, *Trypanoplasma borreli* (data not shown). A rooted phylogenetic tree constructed using nuclear small subunit

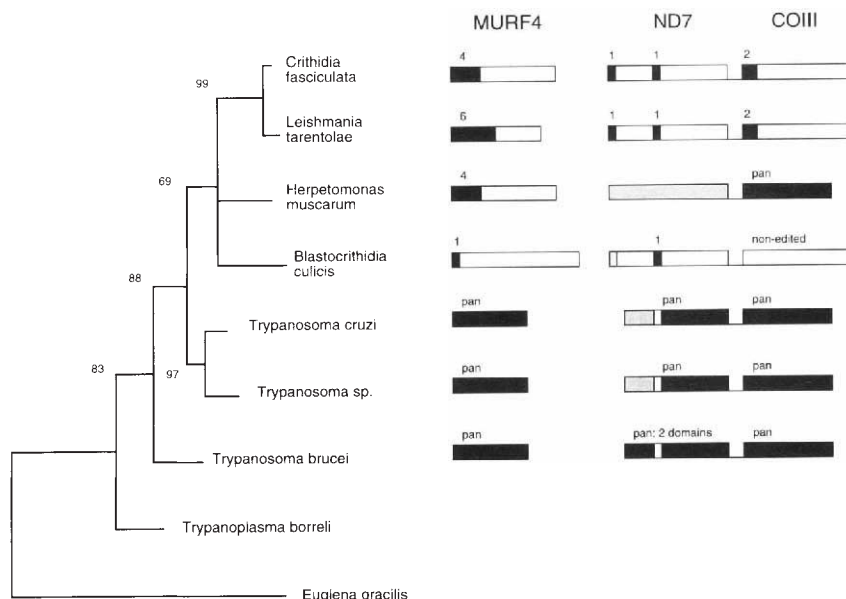
rRNA sequences from monogenetic trypanosomatids (which are parasitic in a single invertebrate host), digenetic trypanosomatids (which parasitize two hosts), and a cryptobiid, is shown in Fig. 1. Using *Euglena gracilis* as an outgroup, the tree is rooted in the *T. borreli* branch. The extreme divergence of the *E. gracilis* sequence from those of the trypanosomatids raises the possibility that this root is artefactually produced by unequal rate effects. However, additional support for this root is provided by the presence of two flagella in both cryptobiids⁴ and euglenoids⁵, and one flagellum in all other taxa. Within the trypanosomatid branch, the most deeply diverging lineage is the salivarian African trypanosome, *Trypanosoma brucei*, followed by the sterocorarian trypanosomes, *T. cruzi* and *T. sp. E1-CP* (a fish parasite). The monogenetic species *Blastocrithidia culicis* and *Herpetomonas muscarum* branch off next, together with a monophyletic clade containing the monogenetic species *Crithidia fasciculata* and the digenetic species *Leishmania tarentolae*.

To investigate the evolution of editing, we studied three mitochondrial genes for which both pan-editing and 5' editing has been reported. These genes encode NADH dehydrogenase subunit 7 (ND7), cytochrome *c* oxidase subunit III (COIII), and maxicircle unidentified reading frame 4 (MURF4), which may represent ATPase subunit 6 (ref. 6).

The MURF4 gene is completely pan-edited in *T. brucei*, and is 5' pan-edited in *L. tarentolae*⁶. We have sequenced the MURF4 genes from the remaining taxa to determine their editing type. The evidence that MURF4 is pan-edited in *T. cruzi* is shown in Fig. 2. An example of a complementary DNA clone of a partially edited RNA amplified by polymerase chain reaction PCR is shown which has an open reading frame (ORF) at the 3' end which translates into a sequence highly similar to the *T. brucei* MURF4 sequence. The upstream sequence represents a misedited sequence, such as those that frequently occur at 'junction regions' in editing intermediates^{7,8}. As the entire DNA

FIG. 1 Phylogeny of kinetoplastid RNA editing.

Aligned sequences of 18S rRNAs from *T. brucei*²⁵, *T. cruzi*²⁶, *L. tarentolae*²⁷, *C. fasciculata*²⁸ and *E. gracilis*²⁹ were retrieved from the Ribosomal Database Project²⁹ and sequences of *T. borreli*, *Trypanosoma* sp. E1-CP, *B. culicis* and *H. muscarum* were obtained by sequencing several independently cloned PCR-amplified gene fragments by standard methods. GenBank accession numbers are L14840 (*T. borreli*), L14841 (*Trypanosoma* sp. E1-CP), L18872 (*H. muscarum*), and U05679 (*B. culicis*). Sequences were aligned using the interactive editor³⁰. The culture of *T. borreli*³¹ was kindly provided by J. Lom. DNA from E1-CP cells was isolated in the Lom laboratory by D.A.M. *H. muscarum* (30261) and *B. culicis* (30268) were obtained from the ATCC. The majority consensus parsimony tree³², which was constructed using 200 bootstrap replicates, is shown. This tree is consistent with the three most parsimonious trees, the first of which (1,174 steps) showed *B. culicis* branching off earliest, followed by *H. muscarum* and then by the *C. fasciculata* and *L. tarentolae* clade. The second tree (2 more steps) made *H. muscarum* and *B. culicis* a monophyletic group, and the third tree (4 more steps) had *H. muscarum* and *B. culicis* reversed from their order in the first tree. The maximum likelihood method³³ (fast DNAmI program provided by G. Olsen) produced a topology identical to that of the second most parsimonious tree. Paralineal distances³⁴, a new algorithm based on very general assumptions, gave the same topology as the tree shown here. Our conclusions on the evolution of editing remain valid with any one of these topologies, and are even stronger for the third most parsimonious tree. The rooted tree shown here indicates that a digenetic life cycle may have evolved independently several times. This differs from our previous speculation that the digenetic life cycle evolved late in the evolution of this group of organisms³⁵ and is consistent with the previous hypothesis of Hoare³⁶. A



diagrammatic representation of the extent of editing in MURF4, ND7 and COIII is shown on the right. Black boxes correspond to pre-edited regions, white boxes to non-edited regions, and grey areas to a lack of information. Numbers above the PERs indicate the actual number of guide RNAs involved (that is, the number of editing blocks) or the estimated minimal number of gRNAs required. GenBank accession numbers are U05816 (*C. fasciculata* MURF4), U05812 (*H. muscarum* MURF4), U05813 (*B. culicis* MURF4), U05879 (*T. cruzi* MURF4), U05818 (*Trypanosoma* sp. E1-CP MURF4), U05817 (*B. culicis* ND7), U05881 (*T. cruzi* ND7), U05815 (*Trypanosoma* sp. E1-CP ND7), U05814 (*B. culicis* COIII), U05878 (*T. cruzi* COIII), U05815 (*Trypanosoma* sp. E1-CP COIII).

FIG. 2 MURF4 is a pan-edited cryptogene in *T. cruzi*. The sequence of one particular cDNA clone of a partially edited mRNA molecule is shown. The portion that corresponds to the fully edited consensus sequence, which was confirmed by sequencing additional clones, underlined. This sequence was aligned with the *T. brucei* edited sequence according to

FIG. 3 Alignment of edited MURF4 sequence from *B. culicis*, *H. muscarmum* and *L. tarentolae*. Nucleotide alignments were made according to amino-acid sequence alignments. *L. tarentolae* editing blocks are indicated below the RNA sequences. Uridines added by editing are shown in lower case, uridines deleted by asterisks. Other designations are as for Fig. 2.

sequence of the PCR-amplified MURF4 fragment from *T. cruzi* is extremely (G + A)-rich (data not shown), as in the case of the pan-edited gene in *T. brucei*, and as editing proceeds 3' to 5', we conclude that MURF4 is completely pan-edited in *T. cruzi*. Sequence analysis of the amplified MURF4 maxicircle fragment from another stercorarian trypanosome, *Trypanosoma* sp. E1-CP, showed that this gene is also pan-edited (data not shown).

ines inserted in 7 sites. The ORF derived from the edited sequence together with the non-edited downstream genomic sequence is similar to that obtained from edited *L. tarentolae* MURF4 sequence, thereby precisely localizing the pre-edited region (PER) in *B. culicis* (Fig. 3). By comparison with *L. tarentolae*, which has a total of six contiguous editing blocks mediated by six overlapping genomic (g)RNAs⁹, only a single gRNA corresponding to the last editing block would be required in *B. culicis* for the observed editing events. A fully edited sequence for MURF4 mRNA from *H. muscarum* was deduced from a collection of overlapping partially edited RNAs. Translation of

L.t. RNA: 5'-(33 nt)-AAAAuuAanGuuuGuu...CGuGuaAuuuuuGuuGGuGUGAGUGUGuuuuuGuuuuuUUUGUCUUVUACCGCUAUU-(795 nt)-UAA

B.c. DNA: (Cyt b)-TAATTAATATGTTTTATTAGAGTAATTACTATAGGAGTAAGTGGTGTTCATATCTTTATCTTTACACGCCATA-(795 nt)-TAA

T.b. RNA: 5'-(35 nt)-uuuuuuuAuGuuuuuuGuuuCGuuGuAuAuuuGuuGGuGuaAGuGGUGuuuuuGuuuuuuuAuCuuuACCuGCCAu-(795 nt)-uAA

L.t. PEP: M F V - R V I F V G V S G V F V : F L S L P A I...

B.c. PEP: M F L F R V I T I G V S G V F I S L S L P A I...

T.b. PEP: M F L F R C I F V G V S G V F V F L S L P A I...

FIG. 4 Comparison of the COIII DNA sequence of *B. culicis* with fully edited mRNA sequences from *L. tarentolae* and *T. brucei*. The edited portions of the sequences are underlined. The initiation and termination

the edited and non-edited sequences allowed a precise localization of the PER (Fig. 3). This editing would require minimally four overlapping gRNAs which would correspond to the gRNAs for the four last editing blocks in *L. tarentolae*. A similar situation was found in the case of the MURF4 gene in *C. fasciculata*, which exhibited a PER four editing blocks in length (data not shown).

The genes for ND7 and COIII are adjacent on the maxicircle. DNA and edited RNA sequences have been reported for *C. fasciculata*¹⁰, *L. tarentolae*¹¹, *T. brucei*^{12,13} and *H. muscarum*¹⁴ (COIII only). The ND7 mRNA in *T. brucei* is pan-edited in two large domains, whereas in *L. tarentolae* and in *C. fasciculata*, the ND7 mRNA is 5' edited in two separate editing blocks which correspond exactly to the 5' termini of the two editing domains in *T. brucei*. Similarly, the two contiguous editing blocks in the *L. tarentolae* COIII mRNA correspond exactly to the 5' terminus of the single large pan-edited domain in *T. brucei*. The COIII gene in *H. muscarum* is pan-edited¹⁴. We have extended this comparative analysis to the phylogenetically significant stercorarian trypanosomes and the *B. culicis* lineages (Fig. 1). Sequence analysis of PCR-amplified maxicircle fragments from *T. cruzi* and *Trypanosoma* sp. EI-CP (data not shown) showed that COIII and ND7 (at least in domain I) are also pan-edited as in *T. brucei*. But in the case of *B. culicis*, the ND7 domain I is represented by a single 5' editing block, as in *L. tarentolae* and *C. fasciculata* (data not shown). On the other hand, the COIII gene in *B. culicis* (Fig. 4) is completely non-edited, as shown by the fact that the translated ORF has a high similarity with COIII protein sequences from other trypanosomatids (69% identity and 84% similarity with the *T. brucei* COIII sequence for example). This is an important finding because it represents the first example of a gene that is present in a completely non-edited version in one species and in a pan-edited version in another species.

The multiple forms of the edited genes for which comparative sequence analysis is available are shown in Fig. 1 on the kinetoplastid phylogenetic tree. Pan-edited MURF4, ND7 and COIII genes are mainly found in the early diverging branches, multiple forms of less edited genes are found in the later separated lineages. The most parsimonious interpretation is that pan-editing of the MURF4 and COIII genes is the primitive state. For COIII, only two independent changes of pan-editing to 5' editing are required: one for the *Blastocrithidia* branch and another for the *Crithidia-Leishmania* clade if pan-editing is ancestral, versus three independent changes from 5' to pan-editing if 5' editing is primitive. For the MURF4 gene, only one change of pan-editing to 5' editing is required. We therefore conclude that pan-editing is a primitive character state that was present in an ancestral trypanosomatid.

Our findings are consistent with the hypothesis that at several times during the course of evolution of the trypanosomatid flagellates, pan-edited genes were replaced with partially edited genes. One possible mechanism is that partially edited RNAs were reverse-transcribed into cDNAs, which were incorporated into the maxicircle genome and replaced the original pan-edited genes^{1,15, 17}. The existence of a non-edited COIII gene in *B. culicis*

codons of COIII in the three species and the termination codon of the upstream cytochrome *b* gene in *B. culicis* are shown in bold. The translated amino-acid sequence is shown below.

indicates that such a mechanism could also completely eliminate editing. This raises the possibility that the ND1, ND4, ND5 and COI genes, which are non-edited in all kinetoplasts examined, may have originally been pan-edited and, in the course of evolution, were replaced by fully edited cDNAs.

A possible driving force for the appearance of lineages in which pan-edited genes have been substituted by partially or completely edited reverse transcripts could be the loss of minicircle-encoded gRNA genes. In *L. tarentolae* each minicircle encodes a single gRNA which only overlaps other gRNAs in the anchor sequence region¹⁸, unlike in *T. brucei*, in which gRNAs overlap even in the guiding region¹⁹. We have shown previously that the relative abundance of different minicircle sequence classes in *L. tarentolae* varies from as few as 30 copies per network of 10,000 minicircles to as many as 2,000 copies⁹. As minicircles apparently segregate randomly between daughter networks, there is a finite chance that a daughter cell will not receive at least one copy of a particular minicircle. A more dramatic loss of one or more minicircle classes could also occur by the phenomenon of 'transkinetoplastidy'²⁰. The subsequent loss of the corresponding gRNA would result in the disruption of an editing cascade. If we make the assumption that a low level of homologous retroposition of partially edited RNAs substituting for pan-edited genes occurs continuously among the 50 or so maxicircle molecules per network, then the cells carrying a retroposed partially edited gene would not require the missing gRNA and would survive. Additional support for this hypothesis is provided by the observation that a loss of minicircle classes is observed in nature (*Trypanosoma equiperdum*)^{21,22} and in the laboratory (*L. tarentolae*-UC strain) (O. Thiemann, D.A.M. and L.S., unpublished results) when cyclical transmission has been prevented. The preservation of pan-editing in cyclically transmitted *T. brucei* is possibly due to the utilization of editing as a translational control mechanism during the complex life cycle of this species²³. The maintenance of an extremely large redundant gRNA repertoire could be an adaptation to prevent the loss of gRNA genes or could indicate that this also is a primitive state.

When and how was RNA editing acquired by the kinetoplastids? Editing could have evolved in the ancestral kinetoplastids by a gene duplication mechanism^{1,24}, or alternatively could have been inherited from the original proto-mitochondrion. To address these questions, an investigation of the possibility of RNA editing in euglenoids and in other mitochondrial-containing protists from the earliest branching lineages, or in eubacteria related to the original endosymbionts, may prove informative. □

Received 3 November 1993; accepted 12 January 1994.

1. Simpson, L., Maslov, D. A. & Blum, B. in *RNA Editing—The Alteration of Protein Coding Sequences of RNA* (ed. Benne, R.) 53–85 (Horwood, New York, 1993).
2. Stuart, K. in *RNA Editing—The Alteration of Protein Coding Sequences of RNA* (ed. Benne, R.) 25–52 (Horwood, New York, 1993).
3. Vickerman, K. in *Biology of the Kinetoplastida* (eds Lumsden, W. H. R. & Evans, D. A.) 1–34 (Academic, London, New York and San Francisco, 1976).
4. Lom, J. in *Biology of the Kinetoplastida* (eds Lumsden, W. H. R. & Evans, D. A.) 267–337 (Academic, London, New York and San Francisco, 1976).
5. Kivic, P. A. & Walne, P. L. *Origins of Life*, 13, 269–288 (1984).
6. Bhat, G. J., Koslowsky, D. J., Feagin, J. E., Smiley, B. L. & Stuart, K. *Cell* **61**, 885–894 (1990).

7. Sturm, N. R., Maslov, D. A., Blum, B. & Simpson, L. *Cell* **70**, 469–476 (1992).
8. Decker, C. J. & Sollner-Webb, B. *Cell* **61**, 1001–1011 (1990).
9. Maslov, D. A. & Simpson, L. *Cell* **70**, 459–467 (1992).
10. Van der Spek, H. et al. *EMBO J.* **10**, 1217–1224 (1991).
11. Blum, B., Bakalar, N. & Simpson, L. *Cell* **60**, 189–198 (1990).
12. Feagin, J. E., Abraham, J. M. & Stuart, K. *Cell* **53**, 413–422 (1988).
13. Koslowsky, D. J., Bhat, G. J., Perrollaz, A. L., Feagin, J. E. & Stuart, K. *Cell* **62**, 901–911 (1990).
14. Landweber, L. F. & Gilbert, W. *Nature* **363**, 179–182 (1993).
15. Landweber, L. F. *BioSystems* **28**, 41–45 (1992).
16. Fink, G. R. *Cell* **49**, 5–6 (1987).
17. Derr, L. D. & Strathern, J. *Nature* **361**, 170–173 (1993).
18. Sturm, N. R. & Simpson, L. *Nucleic Acids Res.* **19**, 6277–6281 (1991).
19. Corell, R. A. et al. *Nucleic Acids Res.* **21**, 4313–4320 (1993).
20. Lee, S. T., Tarn, C. & Chang, K. P. *Molec. Biochem. Parasit.* **58**, 187–203 (1993).
21. Barrois, M., Riou, G. & Galibert, F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3323–3327 (1982).
22. Frasch, A. et al. *Biochim. biophys. Acta* **607**, 397–410 (1980).
23. Koslowsky, D. J., Riley, G. R., Feagin, J. E. & Stuart, K. *Molec. cell. Biol.* **12**, 2043–2049 (1992).
24. Covello, P. S. & Gray, M. W. *Trends Genet.* **9**, 265–268 (1993).
25. Sogin, M., Elwood, H. & Gunderson, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1383–1387 (1986).
26. Hernandez, R., Rios, P., Valdes, a. & Pinero, D. *Molec. Biochem. Parasit.* **41**, 207–212 (1990).
27. Briones, M. R. S. et al. *Molec. Biochem. Parasit.* **53**, 121–128 (1992).
28. Schnare, M., Collings, J. & Gray, M. *Curr. Genet.* **10**, 405–410 (1986).
29. Olsen, G. J. et al. *Nucleic Acids Res.* **20** suppl., 2199–2200 (1992).
30. Olsen, G. J. *Sequence Editor and Analysis Program* (University of Illinois, Urbana, 1990).
31. Peckova, H. & Lom, J. *Parasit. Res.* **76**, 553–558 (1990).
32. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1* (Illinois Natural History Survey, Champaign 1993).
33. Felsenstein, J. *PHYLIP (Phylogeny Inference Package) Version 3.4* (University of Washington, Seattle 1991).
34. Lake, J. A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1455–1459 (1994).
35. Lake, J., de la Cruz, V., Ferreira, P., Morel, C. & Simpson, L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4779–4783 (1988).
36. Hoare, C. A. *The Trypanosomes of Mammals* (Blackwell Scientific, Oxford, 1972).

ACKNOWLEDGEMENTS. We thank G. Olsen and M. Sogin for computer programs and M. Peris for assistance with computer analysis and for discussion. This work was supported in part by grants from the NIH (L.S. and H.A.A.), UNDP/World Bank/WHO/TDR and the Rockefeller Foundation (L.S.), and the NSF (J.A.L.). D.A.M. is on leave from Moscow State University.

Opposite voltage gating polarities of two closely related connexins

Vytas K. Verselis, Christopher S. Ginter & Thaddeus A. Bargiello

Department of Neuroscience, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461-1602, USA

THE molecular mechanisms underlying the voltage dependence of intercellular channels formed by the family of vertebrate gap junction proteins (connexins) are unknown. All vertebrate gap junctions are sensitive to the voltage difference between the cells, defined as the transjunctional voltage, V_j (refs 1, 2), and most appear to gate by the separate actions of their component hemichannels^{3–8}. The heterotypic Cx32/Cx26 junction displays an unpredicted rectification that was reported to represent a novel V_j dependence created by hemichannel interactions, mediated in part by the first extracellular loop E1 (ref. 9). Here we show that aspects of the rectification of Cx32/Cx26 junctions are explained by opposite gating polarities of the component hemichannels, and that the opposite gating polarity of Cx32 and Cx26 results from a charge difference in a single amino-acid residue located at the second position in the N terminus. We also show that charge substitutions at the border of the first transmembrane (M1) and E1 domains can reverse gating polarity and suppress the effects of a charge substitution at the N terminus. We conclude that the combined actions of residues at the N terminus and M1/E1 border form a charge complex that is probably an integral part of the connexin voltage sensor. A consistent correlation between charge substitution and gating polarity indicates that Cx26 and Cx32 voltage sensors are oppositely charged and that both move towards the cytoplasm upon hemichannel closure.

Comparison of E1 sequences of cloned connexins^{10,11} indicates that the first two residues at the proposed M1/E1 border are different in Cx26 (KE) from those of most other members of

the connexin gene family (ES). We exchanged these residues to create Cx32*KE and Cx26*ES and expressed them in *Xenopus* oocytes. The steady-state changes in normalized junctional conductance, G_j , in response to V_j for homotypic wild-type and mutant connexins are illustrated in Fig. 1. Cx32 homotypic junctions gate symmetrically (Fig. 1a), with identical kinetic and steady-state properties about $V_j=0$. Cx26 homotypic junctions gate asymmetrically (Fig. 1d) owing to the presence of an additional dependence on the membrane potential termed V_m , or V_{i-o} (refs 9, 12–14). Best fits of the steady-state data to a Boltzmann relation indicate a calculated gating charge of ~ 2 for Cx32 and ~ 4 for Cx26. Both Cx32*KE (Fig. 1b) and Cx26*ES (Fig. 1e) homotypic junctions retain symmetry about $V_j=0$, but are more sensitive to V_j and have substantially faster kinetics than their wild-type counterparts. Although the G_j-V_j relations of the mutant homotypic junctions, particularly Cx32*KE, have complex shapes that cannot be fitted well by a two-state Boltzmann relation, the maximum slopes in each case are steeper than those of the wild-type junctions, consistent with a substantial increase in the calculated gating charge (see Fig. 1 legend). All the currents decay in a multiexponential fashion (fits to data not shown), indicating, as do the steady-state G_j-V_j relations, that multiple voltage-gated transitions occur in these channels.

In the heterotypic junction Cx26*ES/Cx26 (Fig. 1f), both the kinetic and the steady-state properties are asymmetric about $V_j=0$, resembling Cx26 when the cell expressing Cx26 is made relatively positive, and resembling Cx26*ES when the cell expressing Cx26*ES is made relatively positive. This shows that the component hemichannels retain their characteristics as inferred from the corresponding homotypic junctions, and that Cx26 and Cx26*ES hemichannels close on relative positivity on their cytoplasmic side. The same polarity of closure has been shown for Cx38, Cx37 and Cx40 (refs 4–8). In the corresponding heterotypic junction Cx32*KE/Cx32 (Fig. 1c), the respective kinetic and steady-state hemichannel properties of Cx32 and Cx32*KE are also retained. But the surprising result is that the polarity of V_j sensitivity is opposite to that of Cx26, with Cx32 and Cx32*KE hemichannels closing on relative negativity on their cytoplasmic side (see also refs 15, 16).

To confirm the assignment of positive gating polarity to Cx26 and negative to Cx32, we examined the behaviour of heterotypic Cx32/Cx26, Cx32*KE/Cx26 and Cx32/Cx26*ES junctions. The expected properties of homotypic and heterotypic junctions in which the component hemichannels gate in response to opposite polarities of V_j are shown in Fig. 2a. These expectations were met in each case (Fig. 2b) by the display of a marked asymmetry of the $G-V_j$ relation, with conductance decreasing only for the predicted polarity of V_j . The decrease in G is consistent with the closure of either or both hemichannels, as exemplified by the differences in the resultant $G-V_j$ relations according to properties attributable to the component hemichannels. These results also confirm that gating is predominantly an intrinsic hemichannel property.

Localization of the protein domain responsible for the difference in gating polarity of Cx26 and Cx32 was accomplished by

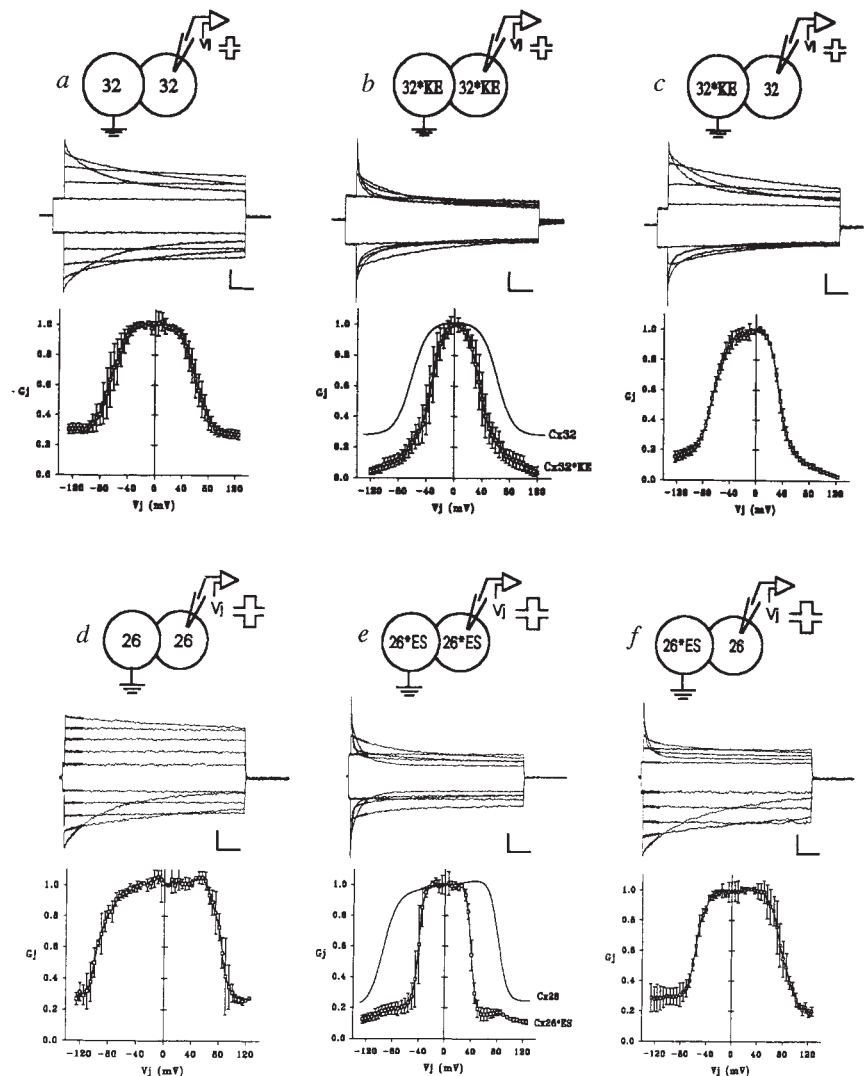
TABLE 1 Summary of gating polarities for point mutations

Mutation	Polarity of closure	Mutation	Polarity of closure
Cx32N2D	+	Cx26D2R	–
Cx32N2E	+	Cx26D2K	–
Cx32N2Q	–	Cx26D2E	+
Cx32N2A	–	Cx32S11D	–
Cx32N2R	–	Cx32Y7D	NE
Cx32N2K	–	Cx32*KE	–
Cx26D2N	–	Cx26*ES	+
Cx26D2Q	–	Cx32*EE	+

Polarity of closure was determined by pairing the mutant connexin with both Cx26 and Cx32 (Fig. 2 legend). NE, no expression.

FIG. 1 KE and ES substitutions markedly alter voltage dependence and reveal opposite gating polarities of Cx32 and Cx26. Normalized steady-state conductance–voltage relations and representative junctional currents, I_j , obtained by dual voltage clamp¹ are shown for homotypic: a, Cx32; b, Cx32*KE; d, Cx26, and e, Cx26*ES junctions, and heterotypic: c, Cx32*KE/Cx32, and f, Cx26*ES/Cx26 junctions. I_j s shown were elicited by V_j s of ± 20 , 40, 60, 80, 100 mV. Depolarization and hyperpolarization of the cell on the right correspond to positive and negative V_j s and downward and upward I_j s, respectively. Both Cx32*KE (b) and Cx26*ES (e) junctions have substantially altered steady-state and kinetic properties compared with their corresponding wild-type junctions (shown as fits (solid line) to a Boltzmann relation) that are maintained in heterotypic Cx32*KE/Cx32 (c) and Cx26*ES/Cx26 (f) junctions. Cx32-like and Cx32*KE-like properties are evident for relative negativity on their cytoplasmic side, whereas Cx26-like and Cx26*ES-like properties are evident on relative positivity. Error bars are standard deviations of means obtained for 6 Cx32, 4 Cx26, 4 Cx26*ES, 4 Cx32*KE, 4 Cx32/Cx32*KE and 3 Cx26*ES/Cx26 cell pairs. Boltzmann parameters for Cx32 are: $A = 0.09 \text{ mV}^{-1}$, $V_0 = 61.6 \text{ mV}$, $G_{\min} = 0.28$, and for Cx26 are: $A = 0.16 \text{ mV}^{-1}$, $V_0 = 95 \text{ mV}$ and $G_{\min} = 0.22$. Only cell pairs displaying $g_j < 5 \mu\text{S}$ were used to avoid problems associated with series access resistance^{24,25}. Time calibration bars are 2 seconds. Current calibration bars are: 70 nA (Cx32); 100 nA (Cx32*KE); 100 nA (Cx32*KE/Cx32); 70 nA (Cx26); 100 nA (Cx26*ES); 40 nA (Cx26*ES/Cx26).

METHODS. A family of junctional currents for each cell pair was generated by applying V_j s of $\pm 5 \text{ mV}$ to $\pm 125 \text{ mV}$ in 5-mV increments with 90-s interpulse intervals. Steady-state and initial g_j was obtained by fitting I_j to exponentials of the form $A(1)e^{-V_j/\tau_1} + \dots + A(n)e^{-V_j/\tau_n} + C$, where C is the steady-state current and $A(n)$ are initial current amplitudes of the component exponentials. Typically, 2–3 exponentials were necessary to fit the data. Each V_j step was preceded by a small prepulse (10 mV) of short duration (100–500 ms) to assess the level of expression, and was used to normalize I_j within a family of currents; g_j was normalized in each cell pair to the value at $V_j = 0$, obtained from the slope of the linear part of the I - V curve about $V_j = 0$. Apparent gating charge was obtained when possible by fitting G_j to a



Boltzmann relation of the form described in ref. 9. Although the existence of multiple states strictly precludes the application of a two-state Boltzmann equilibrium, comparison of fits to such a Boltzmann provide useful indexes for relative changes in gating charge²⁶. Preparation of oocytes, recording solutions, synthesis and injection of RNA and methods of mutagenesis have all been described^{9,27}.

the construction of a series of chimaeras in which segments of Cx32 were substituted into Cx26 (Fig. 3). All chimaeras that contain the N terminus (NT) of Cx32 were found to have Cx32 polarity of closure. Most of these display less steep G_j - V_j relations, as exemplified by Cx26*32NTM1/Cx32 junctions (Fig. 3b). A G_j - V_j relation indistinguishable from wild-type Cx32 is observed in Cx26*32NT-M2 hemichannels (not shown), suggesting that the sequence beyond M2 has no role or structural equivalence in the expression of Cx32 voltage-dependence.

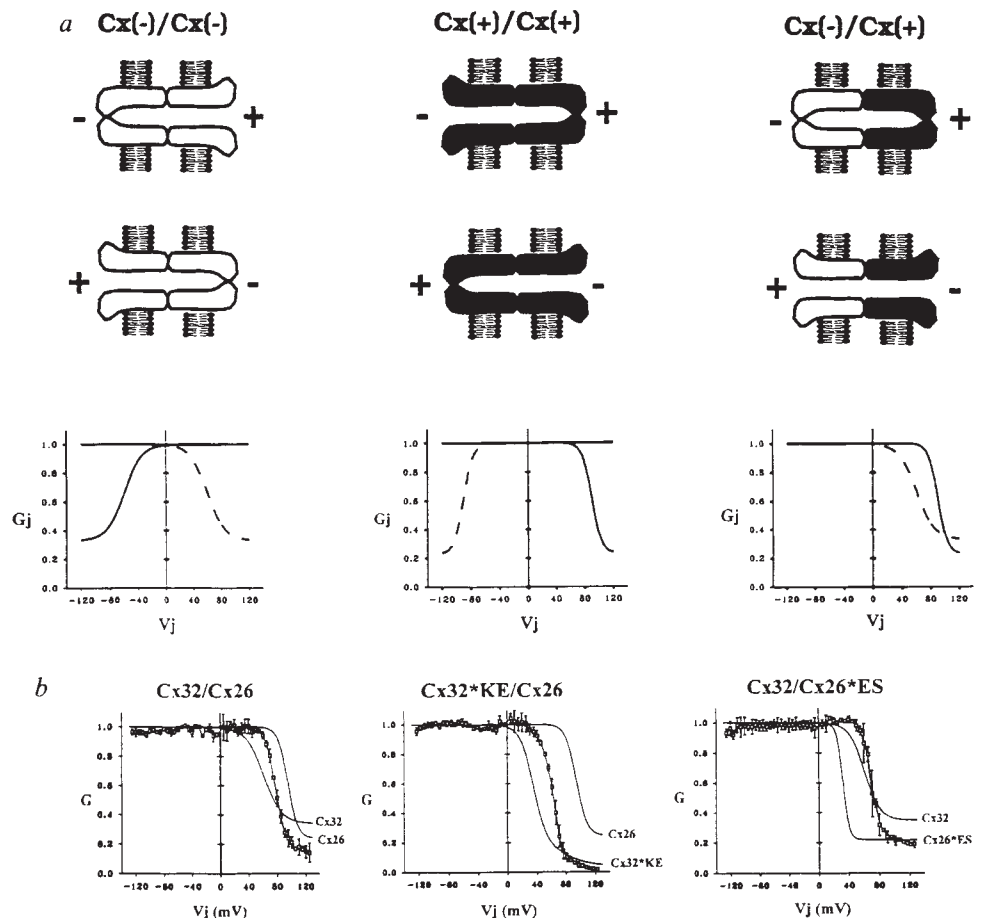
In the Cx26 32NT-V13 and Cx26*32NTM1 chimaeras, 35 of the 36 charged amino acids present in Cx26 are conserved. The only difference in charge resides at position 2 in the N-terminal domain, where a negatively charged aspartic acid (D) in Cx26 is replaced with a neutral asparagine (N) in Cx32. The point mutation, N2D, reverses the gating polarity of Cx26*32NT-V13, Cx26*32NTM1, and wild-type Cx32 hemichannels. Table 1 summarizes the gating polarity of amino-acid substitutions at position 2 in Cx32 and Cx26. In all cases the polarity of hemichannel closure correlates with the charge of the amino-acid residue at position 2, indicating a dominant electrostatic effect on voltage gating. The presence or absence of a negative charge in each of

the six connexin subunits within a hemichannel can account for the difference in gating polarity and calculated gating charge of Cx26 (~ 4) and Cx32 (~ 2). Each Cx26 subunit would contribute 0.66 negative charges to the voltage sensor, whereas each Cx32 subunit would contribute 0.33 positive charges. The presence of a negative charge in the N terminus of each Cx32 subunit would be more than sufficient to reverse the polarity of its voltage sensor, as would the removal of the same charge from Cx26.

We examined the effect of the charge substitution within the N terminus of Cx32 by moving the charged residue to positions 7 and 11, corresponding to positions where negatively charged residues reside in other connexins. Hemichannels formed by Cx32S11D do not reverse Cx32 gating polarity (Table 1). We did not observe expression of Cx32Y7D channels. Thus, the electrostatic effect is likely to involve residues proximal to position 11 in Cx32.

The maintenance of polarity and the apparent increase in calculated gating charge of Cx26*ES and Cx32*KE hemichannels also correlates with a charge effect that makes the Cx26 voltage sensor relatively more negative (ES for KE) and the Cx32 voltage sensor relatively more positive (KE for ES). The addition

FIG. 2 a, Expected G_j - V_j relations and diagrammatic representations of gap junction hemichannel open/closed configurations at large V_j s of either sign for homotypic junctions in which the component hemichannels gate on negativity, $Cx(-)$, and positivity, $Cx(+)$, relative to their cytoplasmic ends and a heterotypic junction composed of $Cx(-)$ and $Cx(+)$ hemichannels. The solid and dashed lines represent Boltzmann relations describing the gating of hemichannels on the right and left side of the junction, respectively. Homotypic junctions maintain symmetry about $V_j = 0$, but opposite hemichannels close for a given polarity of V_j . The G_j - V_j relation of the heterotypic $Cx(-)/Cx(+)$ junction is asymmetric, because both hemichannels close for V_j which is both relatively positive to $Cx(+)$ and relatively negative to $Cx(-)$. Neither hemichannel closes for the opposite polarity of V_j . b, Conductance-voltage relations for three heterotypic junctions, $Cx32/Cx26$, $Cx32/Cx26^*ES$ and $Cx32^*KE/Cx26$. Because steady-state changes in G_j in these junctions are the result of fast and slow processes^{9,12}, the data and fits are plotted as the ratio of steady-state conductance, G_j , to initial conductance, G_0 . The ratio is termed G and is used only to illustrate changes in the slow process to which these data on polarity reversal apply. Independence of fast and slow processes is not implied. The mechanism of the fast processes in $Cx32/Cx26$ junctions remains unknown. The G - V_j plots of the component hemichannels, solid lines, represent Boltzmann relations obtained using parameters fit to the ratio, G , for homotypic junctions of each



type. Error bars are standard deviations of means obtained for 5 $Cx32/Cx26$, 3 $Cx32^*KE/Cx26$ and 3 $Cx32/Cx26^*ES$ cell pairs. Records and analyses were obtained as described for Fig. 1.

of a negative charge at position 42, $Cx32^*EE$, is sufficient to reverse the gating polarity of $Cx32$. We also observed that $Cx32^*KE$ suppresses the ability of $Cx32N2D$ to reverse the polarity of $Cx32$ closure, consistent with addition of a positive charge (data not shown).

The simplest picture that emerges from our results is that amino acids at or near the N terminus and the M1/E1 border form part of a charge complex that is an integral part of the voltage sensor. Based on the accepted membrane topology of connexins¹⁰ (see also Fig. 3b), it is likely that the amino terminus is located within or close to the channel pore where it can sense the transjunctional field. Given this topology, the unblocked N-terminal methionine residue would contribute a positive charge to this complex. If this charge complex does not move in the electric field, it must determine the movement of other charges that represent the 'formal' voltage sensor. In either case, conserved charges in M1, M3 and at the border of other transmembrane domains are good candidates to be additional components of the voltage sensor. The consistent pattern of electrostatic effects described indicates that the $Cx26$ and $Cx32$ voltage sensors are oppositely charged, with $Cx32$ positive and $Cx26$ negative. Both sensors would move towards the cytoplasmic side of the hemichannel upon closure, implying a conservation of the gating mechanism in these connexins. As none of the point mutations that reverse polarity of closure exactly recreates wild-type voltage dependence, additional structural changes, either induced by the mutations or inherent in $Cx26$ or $Cx32$, alter the calculated gating charge. However, these structural changes have

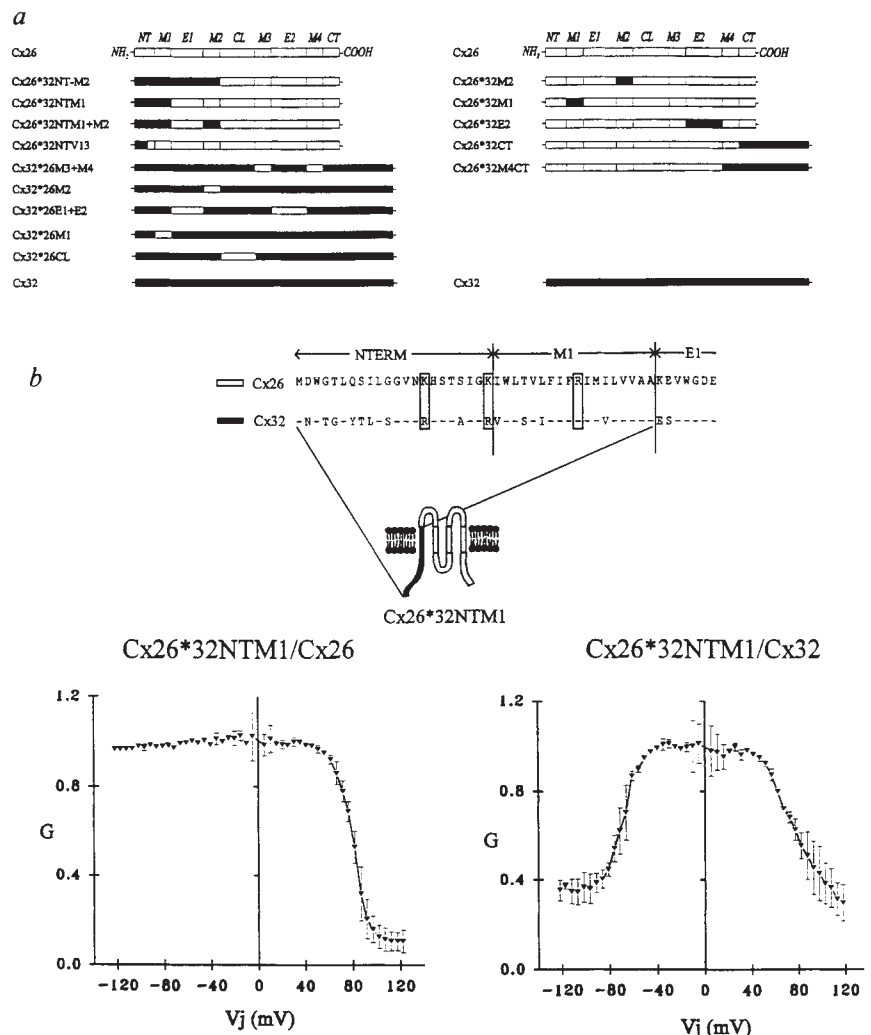
a secondary role to the electrostatic effects reported here.

It has been reported¹⁷ that substitution of a conserved proline residue (P87) located within M2 of $Cx26$ reversed gating polarity and it was concluded that this residue functions as a 'transduction element' in connexin voltage-dependent gating. We observe no such effect with comparable mutations in $Cx32$. Notably, the behaviour of the $Cx32N2D$ mutation is unaffected by the $Cx32P87G$ substitution (Y. Ri, V.K.V. and T.A.B., unpublished observations). Therefore, the mechanism of $Cx32$ polarity reversal described here cannot explain the different effect of P87G substitutions in $Cx26$ and $Cx32$. P87G mutations may induce structural changes in some connexins that perturb the electro-

Received 5 August 1993; accepted 18 January 1994.

1. Spray, D. C., Harris, A. L. & Bennett, M. V. L. *J. gen. Physiol.* **77**, 77-93 (1981).
2. Bennett, M. V. L. & Verselis, V. K. *Semin. Cell Biol.* **3**, 29-47 (1992).
3. Harris, A. L., Spray, D. C. & Bennett, M. V. L. *J. gen. Physiol.* **77**, 95-117 (1981).
4. Swenson, K. I., Jordan, J. R., Beyer, E. C. & Paul, D. L. *Cell* **57**, 145-155 (1989).
5. Werner, R., Levine, E., Rabadán-Diehl, C. & Dahl, G. *Proc. R. Soc. Lond.* **243**, 5-11 (1991).
6. Hennemann, H. et al. *J. Cell Biol.* **117**, 1299-1310 (1992).
7. Bruzzone, R., Haefliger, J. A., Gimlich, R. L. & Paul, D. L. *Molec. Biol. Cell* **4**, 7-20 (1993).
8. Bennett, M. V. L., Verselis, V., White, R. L. & Spray, D. C. in *Gap Junctions* (eds Hertzberg, E. L. & Johnson, R. G.) 287-304 (Liss, New York, 1988).
9. Rubin, J. B., Verselis, V. K., Bennett, M. V. L. & Bargiello, T. A. *Biophys. J.* **62**, 197-209 (1992).
10. Bennett, M. V. L. et al. *Neuron* **6**, 305-320 (1991).
11. Haefliger, J. et al. *J. biol. Chem.* **267**, 2057-2064 (1992).
12. Barrio, L. C. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8410-8414 (1991).
13. Obaid, A. L., Socioar, S. J. & Rose, B. J. *Membr. Biol.* **73**, 69-89 (1983).
14. Verselis, V. K., Bennett, M. V. L. & Bargiello, T. A. *Biophys. J.* **59**, 114-126 (1991).
15. Paul, D. L., Ebihara, L., Takemoto, L. J., Swenson, K. I. & Goodenough, D. A. *J. Cell Biol.* **115**, 1077-1089 (1991).

FIG. 3 Reversal of Cx26 gating polarity is evident in chimaeras in which the proximal N-terminal segment of Cx26 is replaced by Cx32 sequence. **a**, Shown schematically is a series of Cx32 and Cx26 chimaeras segmented according to the accepted membrane topology of connexins¹⁰. Cx26 sequence is indicated by open boxes. Chimaeras on the left close on relative negativity on their cytoplasmic ends; those on the right on relative positivity. **b**, Primary sequences of Cx26 and Cx32 from the N terminus through to the proximal portion of E1 are shown with a schematic of the putative connexin membrane topology and $G-V_j$ relations of chimaeric Cx26*32NTM1/Cx26 and Cx26*32NTM1/Cx32 heterotypic junctions. Reductions in G (Fig. 2 legend) for a single polarity in Cx26*32NTM1/Cx26 junctions and for both polarities in Cx26*32NTM1/Cx32 junctions indicates that component hemichannels composed of Cx26*32NTM1 close on relative negativity, as do hemichannels composed of Cx32. The steady-state properties for Cx26*32NTM1 (right side of Cx26*32NTM1/Cx32 $G-V_j$ relation) is less steep than for Cx32 and does not reach a plateau by 120 mV. Boltzmann parameters for relative negativity on the Cx32 side are: $A=0.126\text{ mV}^{-1}$, $V_0=68.7\text{ mV}$, $G_{\min}=0.35$; and for relative negativity on the Cx26*32NTM1 side are: $A=0.079\text{ mV}^{-1}$, $V_0=74.6\text{ mV}$ and $G_{\min}=0.31$. Error bars are standard deviations of means obtained for 3 cell pairs of each heterotypic junction.



static mechanism proposed here, or alternatively create a novel gating mechanism.

Together with connexins, the gene family encoding inward-rectifying K^+ channels appear to have members that display intrinsic voltage-dependencies of opposite polarities^{18–20}. Inward rectifiers have little sequence homology to the other voltage-gated K^+ channels, notably they lack the S4 domain which is an integral part of the voltage sensor^{21,22}. The mitochondrial voltage-dependent anion channel VDAC is believed to be a monomer which is capable of gating to low-conductance states for either polarity of membrane voltage²³. It will be of interest to compare with connexins the molecular basis of polarity determination in these channels. □

- Devries, S. H. & Schwartz, E. A. *J. Physiol.* **455**, 201–230 (1992).
- Suchyna, T. M., Lie, X. X., Gao, F., Fournier, C. R. & Nicholson, B. J. *Nature* **365**, 847–849 (1993).
- Ho, K. et al. *Nature* **362**, 31–38 (1993).
- Kubo, Y., Baldwin, T. J., Jan, Y. N. & Jan, L. Y. *Nature* **362**, 127–133 (1993).
- Silver, M. R. & DeCoursey, T. E. *J. gen. Physiol.* **96**, 109–134 (1990).
- Noda, M. et al. *Nature* **312**, 121–127 (1984).
- Liman, E. R., Hess, P., Weaver, F. & Koren, G. *Nature* **353**, 752–756 (1991).
- Thomas, L., Blachly-Dyson, E., Colombini, M. & Forte, M. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5446–5449 (1993).
- Rubin, J. B. thesis, Albert Einstein Coll. Med., New York (1992).
- Wilders, R. & Jongsma, H. J. *Biophys. J.* **63**, 942–953 (1992).
- Papazian, D. M., Timpe, L. C., Jan, Y. N. & Jan, L. Y. *Nature* **349**, 305–310 (1991).
- Rubin, J. B., Verselis, V. K., Bennett, M. V. L. & Bargiello, T. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3820–3824 (1992).

ACKNOWLEDGEMENTS. We thank J. Rubin, A. Finkelstein, J. Ballasteros, H. Weinstein and M.V.L. Bennett for discussion and suggestions, M. Chanson and E. Herzberg for comments on the manuscript, and A. Weiss and Y. Ri for assistance. This study was funded by the NIH.

Histone H1 is located in the interior of the chromatin 30-nm filament

V. Graziano, S. E. Gerchman, D. K. Schnelder & V. Ramakrishnan*

Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

THE linker histone H1 binds to the nucleosome and is essential for the organization of nucleosomes into the 30-nm filament of chromatin. It has been implicated in the repression of transcription^{2–5}, and phosphorylation of H1 may be involved in cell-cycle-dependent chromatin condensation and decondensation⁶. A long-standing issue concerns the location of H1 in the chromatin filament⁷. The original solenoidal model⁸ proposes that H1 is inside the 30-nm filament, but other models, also helical, suggest a variable⁹ or more accessible¹⁰ location for H1. Investigations to determine the location of the linker histone based on its accessibility to antibodies^{11–15} or immobilized proteases¹⁶ under various ionic conditions have yielded conflicting results. Here we use neutron scattering in a direct structural determination to show that H1 is located in the interior of the filament.

* To whom correspondence should be addressed.

Small-angle neutron scattering has been used to measure the mass per unit length, M_l , and cross-sectional radius of gyration, R_x , of chromatin 30-nm filaments^{17,18}. Such scattering arises from the contrast, or the difference $\Delta\rho$, between the mean scattering length density ρ_p of the macromolecule and ρ_s of the solvent. For neutron scattering, contrast can be changed in two ways: by altering the H_2O/D_2O ratio in the buffer (ρ_s), and by deuterating the macromolecule (ρ_p). By deuterating H1, its contrast can be enhanced relative to the rest of the chromatin, and information obtained directly about its radial location in the filament. We describe the results of contrast variation on three types of chromatin 30-nm filament: native chromatin (N-chromatin), chromatin reconstituted with unlabelled H1 (H-chromatin), and chromatin reconstituted with deuterated H1 (D-chromatin).

Figure 1a shows cross-sectional Guinier plots of N- and H-chromatin in 30 and 80% D_2O ; Fig. 1b shows plots for D-chromatin under the same conditions. There is little change in

the slope of a linear fit to the low-angle data for N- or H-chromatin at these two contrasts. Moreover, the curves for N- and H-chromatin are superimposable. However, there is a significant change in the slope at the two contrasts for D-chromatin. In Table 1 we show the values of the intercept $[qI(q)]_0$ and R_x for various percentages of D_2O in the buffer. A plot of the forward scattering amplitude $[qI(q)]_0^{1/2}$ versus per cent D_2O in the solvent yields the match point at which the mean contrast of the filament is zero (Fig. 2a). The data for the N- and H-chromatin completely overlap. As expected, the match point is shifted to a higher density for D-chromatin; however, the slopes for all three chromatin samples are parallel, showing that they have identical particle volumes. The match points are in good agreement with the values calculated on the basis of chemical composition. A mass per unit length, M_l , of 6 nucleosomes per 11 nm was obtained for all three samples of chromatin from the value of $[qI(q)]_0$ in H_2O .

The variation of R_x^2 with inverse contrast is shown in Fig. 2b.

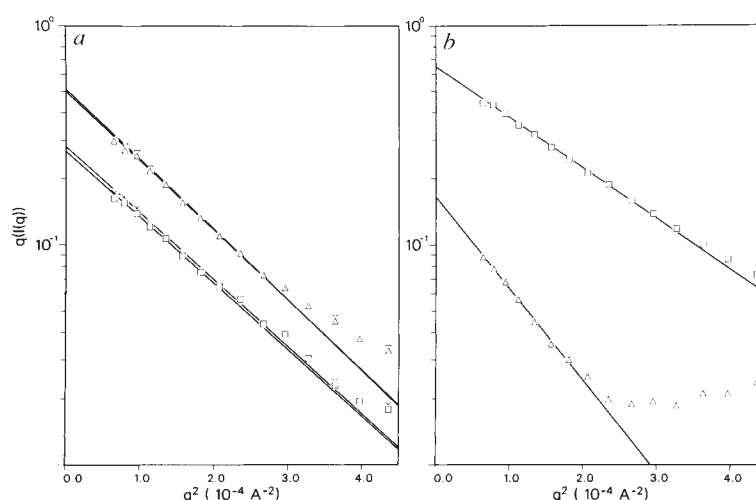
TABLE 1 Forward scatter $[qI(q)]_0$ and cross-sectional radius of gyration R_x at various contrast points for three types of chromatin sample

%D ₂ O	N-chromatin		H-chromatin		D-chromatin	
	$[qI(q)]_0$	R_x (Å)	$[qI(q)]_0$	R_x (Å)	$[qI(q)]_0$	R_x (Å)
0	1.384 ± 0.013	115.6 ± 0.5	1.329 ± 0.016	117.7 ± 0.6	2.061 ± 0.011	109.5 ± 0.3
20	0.582 ± 0.004	116.5 ± 0.5	0.562 ± 0.005	118.8 ± 0.6	1.018 ± 0.005	106.6 ± 0.3
30	0.266 ± 0.003	117.7 ± 0.9	0.280 ± 0.003	118.3 ± 0.7	0.650 ± 0.003	103.2 ± 0.3
40	0.090 ± 0.002	116.9 ± 1.5	0.088 ± 0.004	122.8 ± 2.4	0.316 ± 0.002	98.5 ± 0.5
60	—	—	0.047 ± 0.002	121.5 ± 2.5	—	—
75	0.345 ± 0.003	119.5 ± 0.5	0.334 ± 0.004	119.4 ± 0.6	—	—
80	0.499 ± 0.004	120.8 ± 0.4	0.512 ± 0.005	121.5 ± 0.4	0.165 ± 0.004	138.5 ± 1.4
85	0.688 ± 0.005	118.3 ± 0.4	0.676 ± 0.005	119.7 ± 0.4	0.275 ± 0.004	137.2 ± 0.8
90	0.920 ± 0.006	120.1 ± 0.3	0.868 ± 0.006	120.1 ± 0.3	0.405 ± 0.006	135.7 ± 0.8
100	1.437 ± 0.005	119.2 ± 0.2	1.264 ± 0.014	120.7 ± 0.6	0.693 ± 0.007	130.9 ± 0.6

Values obtained were from linear fits to Guinier plots such as those shown in Fig. 1. The per cent D_2O values are nominal ones that were used for mixing appropriate ratios of the H_2O and D_2O samples. The exact solvent density for these samples was calculated from their transmissions, and these were used in subsequent analyses. The errors shown are those propagated from counting statistics and are underestimates of the true errors. The values for the forward scatter and radius of gyration are nearly the same for N- and H-chromatin, and the radius of gyration changes only slightly with contrast. On the other hand, there is a 40 Å change in R_x for D-chromatin over the contrast range studied. The units for the forward scatter are arbitrary. The forward scatter was also converted to an absolute scale using nucleosome core particles (data not shown) as a standard to calculate the mass per unit length of the various chromatin samples²⁵.

FIG. 1 Cross-sectional Guinier plots for chromatin. Samples of N-, H- and D-chromatin (see text for description of notation) were used in small-angle neutron-scattering experiments. a, Data for N- and H-chromatin. Δ , N-chromatin in 80% D_2O ; $\square$, N-chromatin in 30% D_2O ; ∇ , H-chromatin in 80% D_2O ; $\times$, H-chromatin in 30% D_2O . b, D-chromatin in 80% (Δ) and 30% ($\square$) D_2O .

METHODS. Samples were made as described^{18,21,25}, and in all cases consisted of chromatin fragments with DNA lengths in excess of 15 kilobases. The T7 expression system²⁶ was used to overproduce unlabelled H1 and deuterated H1 in *E. coli*. Deuterated H1 was produced by growth and induction of *E. coli* in minimal medium containing deuterated succinate as the carbon source. The gene used was chicken H1-11L, which corresponds to the predominant H1 in chicken erythrocytes²⁷. Samples were dialysed into 85 mM NaCl, 5 mM HEPES, 0.2 mM EDTA, apparent pH 7.5, at a concentration of ~ 8 –10 mg ml^{-1} of chromatin in H_2O and D_2O . We have determined previously^{18,21} that these chromatin concentrations produce nearly ideal values for the scattering curves, and that chromatin is fully condensed at this salt concentration. We have previously shown that the limiting mass per unit length is the same regardless of whether NaCl or $MgCl_2$ is used to condense chromatin¹⁸. Intermediate H_2O/D_2O ratios were obtained by mixing appropriate volumes of the H_2O and D_2O samples. Neutron scattering was done on beam-line H9B of the High Flux Beam Reactor at Brookhaven. Neutrons of wavelength 7.4 Å from the cold moderator were used. Data were collected on a 2-D He^3 multi-wire detector, radially integrated, corrected for background and normalized for beam intensity,

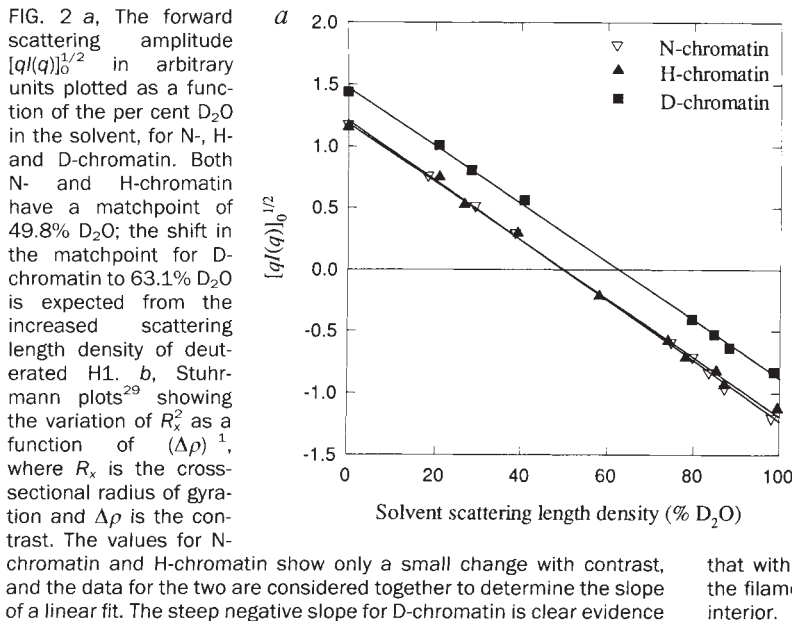


thickness, sample transmission and concentration as before^{18,21} to yield the scattered intensity $I(q)$ as a function of the scattering vector q , given by $4\pi \sin \theta / \lambda$ (where λ is the wavelength of the radiation and 2θ is the scattering angle). The low-angle part of plots of $\ln [qI(q)]$ versus q^2 were fitted with a straight line to obtain the forward scatter $[qI(q)]_0$ and the cross-sectional radius of gyration R_x from the intercept and slope respectively²⁸. The y-axis is in arbitrary units.

For N- and H-chromatin, R_x^2 shows little change with contrast; the small negative slope is evidence that the interior of the filament is slightly DNA-rich, implying that the linker DNA is inside. But the steep negative slope of the data from D-chromatin is evidence that the increased scattering-length density from deuteration, which comes from H1, lies in the interior of the filament. The line for N- and H-chromatin intersects with that

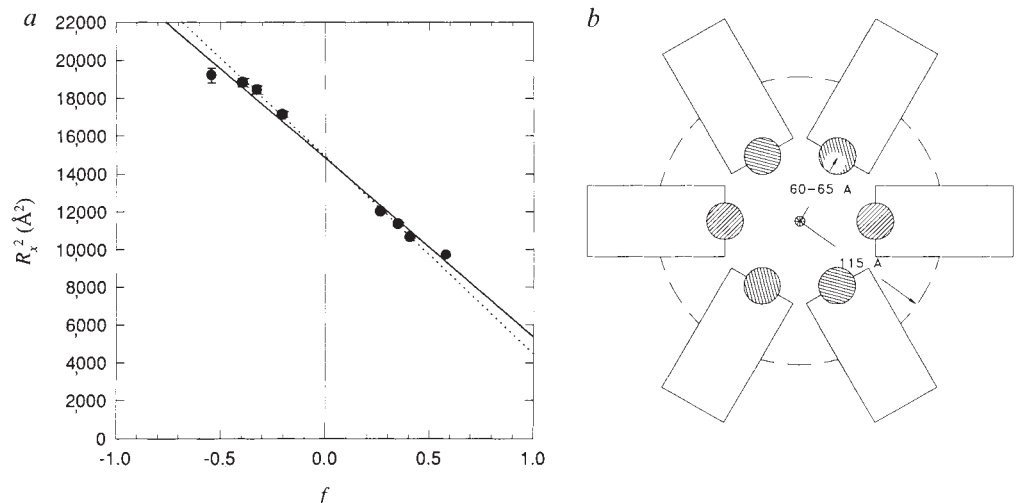
for D-chromatin at the infinite-contrast point ($1/\Delta\rho=0$). This is evidence that all three samples have the same shape, and differ only in the relative contrasts of the various components.

In Fig. 3a, the variation of R_x^2 is shown as a function of the fractional contribution, f , that H1 makes to the scattering¹⁹. From this it is possible to estimate the values of R_x for H1, corresponding to $f=1$, and the remainder of chromatin, corre-



that with the deuteration of H1 there is higher density in the core of the filament than in the periphery²⁹, showing that H1 is located in the interior.

FIG. 3 a, A plot of the square of the cross-sectional radius of gyration of D-chromatin as a function of the fractional mean excess scattering, f , of the deuterated H1. The fraction f that H1 contributes to the scattering is the ratio of $(\Delta\rho)V$ for H1 to that for the whole chromatin, where $\Delta\rho$ is the contrast and V is the volume. This fraction changes with the solvent scattering length density. Thus, as the overall contrast changes, the relative contribution of deuterated H1 to the scattering changes. If R_{H1} is the cross-sectional radius of gyration of the H1 component, R_{ch} is the cross-sectional radius of gyration of the remainder of chromatin, and f is the fraction of the scattering that comes from H1, then the measured cross-sectional radius of gyration R_x is given by $R_x^2 = fR_{H1}^2 + (1-f)R_{ch}^2 + f(1-f)\Delta^2$, where Δ is the separation between the centres of mass of the H1 and remaining components. With chromatin, one can assume that, on average, H1 and the nucleosome are distributed concentrically about the fibre axis, so that Δ is approximately zero. This assumption is true for the various helical models for chromatin. It is also borne out by the linearity of the data in Fig. 2b and in this figure. With this assumption, the relation $R_x^2 = fR_{H1}^2 + (1-f)R_{ch}^2$ holds. Thus a linear fit to the data allows the estimation of R_{H1} and R_{ch} from the intercepts at $f=1$ and $f=0$ respectively. From the data, we estimate that $R_{ch}=122$ Å; R_{H1} is 73 Å if all points are included (solid line), and 67 Å if the two outermost points, corresponding to low-contrast measurements, are ignored in the fit (dotted line). b, Diagram showing the radial location of H1 and the nucleosome in the 30-nm filament of chromatin, based on the data in a. For an object with a radius of gyration R_g about its own centre of mass, and a measured radius of gyration R_x about a filament axis, the distance R of the centre of mass from the



filament axis obeys the parallel axis theorem: $R_x^2 = R^2 + R_g^2$. For H1, we use $R_g=30$ Å, which we obtained from measurements of free H1 in solution (data not shown). For the remainder of chromatin, we assume that most of the scattering comes from the nucleosome, for with R_g is about 45 Å. This results in distances from the fibre axis of 60–65 Å and 115 Å respectively for the centres of mass of H1 and the nucleosome. The distances shown have been rounded off to the nearest 5 Å, and are likely to be no more accurate than 10 per cent. These values are not very sensitive to the choice of the R_g chosen for H1 and the nucleosome. The figure shows a view looking down the fibre axis, and the hatched circle showing H1 has the approximate dimension of the globular domain of H1. The nucleosome is represented by a box of dimensions of 110 Å by 57 Å. The inner face of the nucleosome thus falls at approximately the same radial location as the H1 centre of mass, suggesting that H1 binds to the face of the nucleosome that is presented to the interior of the filament.

sponding to $f=0$. For the remainder of chromatin, this value is 122 Å, whereas for H1 it is 73 Å if all points are included, and 67 Å if the two outermost low-contrast points are omitted from the fit. If we assume that the scattering from the remainder of chromatin comes mainly from the nucleosome, we can apply the parallel-axis theorem to estimate the distance of the centres of mass of H1 and the nucleosome from the filament axis (Fig. 3 legend). A model based on the results of this analysis is shown in Fig. 3b. The model places H1 at about the same radial location as the inner face of the nucleosome, suggesting that H1 interacts with this face in the filament. Our model also makes it unlikely that the globular domains of H1 interact directly with each other in the filament, but possibly the extended tails make contact. The model is also compatible with a scattering study on short chromatin fragments²⁰ which showed that the interior was filled rather than hollow.

We have previously considered in detail the nature of the reconstituted samples used here²¹. By every criterion tested, the reconstituted material is structurally indistinguishable from native chromatin: both reconstituted and native chromatin have a mass per unit length of ~6 nucleosomes per 11 nm in the condensed state. The rate of digestion of chromatin into mononucleosomes by micrococcal nuclease, observation of the 166-base-pair chromatosomal pause during digestion by nuclease, preservation of the 11-nm repeat of chicken erythrocyte chromatin, and the quantitative behaviour of the mass per unit length and radius of gyration during condensation, are all identical for native and reconstituted chromatin. Further, as we show here, the scattering from native and reconstituted chromatin is identical even as a function of contrast. Finally, it has been shown²² that the reassociation of linker histone with the nucleosome is specific. Thus we have every reason to believe that the samples described here are representative of native chromatin filaments.

It has been observed that H1 can exchange readily between filaments^{23,24}, suggesting that chromatin is a dynamic structure; thus H1 may become accessible during the course of an experiment and react irreversibly with antibodies, and this may be a reason for the conflicting interpretations of previous studies. Evidence of this comes from Cattini and Allan¹³, who found a difference in accessibility between H1 that had been crosslinked and H1 that had not. We have measured time-averaged structural parameters that are directly related to the distribution of density in the filament and which are therefore not affected by transient changes in H1 location.

In conclusion, our data show that chromatin 30-nm filaments have ~6 nucleosomes per 11 nm, with histone H1 at about the same radial location as the inner face of the nucleosome. This picture is in agreement with the solenoidal model as originally proposed, and the fact that H1 is known to interact with the nucleosome at the entry and exit points of nucleosomal DNA. □

Received 4 November 1993; accepted 24 January 1994.

- Thoma, F., Koller, T. & Klug, A. *J. Cell Biol.* **83**, 403–427 (1979).
- Schliessel, M. S. & Brown, D. D. *Cell* **37**, 903–913 (1984).
- Wolffe, A. P. & Brown, D. D. *Science* **241**, 1626–1632 (1988).
- Kamakaka, R. T. & Thomas, J. O. *EMBO J.* **9**, 3997–4006 (1990).
- Croston, G. E., Kerrigan, L. A., Lira, L. M., Marshak, D. R. & Kadonaga, J. T. *Science* **251**, 643–649 (1991).
- Roth, S. Y. & Allis, C. D. *Trends biochem. Sci.* **17**, 93–98 (1992).
- Felsenfeld, G. & McGhee, J. D. *Cell* **44**, 375–377 (1986).
- Finch, J. T. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1897–1901 (1976).
- McGhee, J. D., Nickol, J. D., Felsenfeld, G. & Rau, D. C. *Cell* **33**, 831–841 (1983).
- Woodcock, C. L. F., Frado, L.-L. & Rattner, J. B. *J. Cell Biol.* **99**, 42–52 (1984).
- Dimitrov, S. I., Russanova, V. R. & Pashev, I. G. *EMBO J.* **6**, 2387–2392 (1987).
- Russanova, V. R., Dimitrov, S. I., Makarov, V. L. & Pashev, I. G. *Eur. J. Biochem.* **167**, 321–326 (1987).
- Cattini, P. A. & Allan, J. J. *Histochem. Cytochem.* **36**, 425–432 (1988).
- Thibodeau, A. & Ruiz-Carrillo, A. *J. Biol. Chem.* **263**, 16236–16241 (1988).
- Banchev, T. B., Srebeva, L. N. & Zlatanova, J. S. *Molec. cell. Biochem.* **95**, 167–175 (1990).
- Leuba, S. H., Zlatanova, J. & van Holde, K. J. *molec. Biol.* **229**, 917–929 (1993).
- Suau, P., Bradbury, E. M. & Baldwin, J. P. *Eur. J. Biochem.* **97**, 593–602 (1979).
- Gerchman, S. E. & Ramakrishnan, V. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7802–7806 (1987).
- Serdyuk, I. N. *Brookhaven Symp. Biol.* **27**, IV:49–60 (1975).
- Lasters, I. et al. *Eur. J. Biochem.* **151**, 283–289 (1985).
- Graziano, V., Gerchman, S. E. & Ramakrishnan, V. *J. molec. Biol.* **203**, 997–1007 (1988).
- Hayes, J. J. & Wolffe, A. P. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6415–6419 (1993).

- Caron, F. & Thomas, J. O. *J. molec. Biol.* **146**, 513–537 (1981).
- Thomas, J. O. & Rees, C. *Eur. J. Biochem.* **134**, 109–115 (1983).
- Graziano, V. & Ramakrishnan, V. *J. molec. Biol.* **214**, 897–910 (1990).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185**, 61–89 (1990).
- Shannon, M. F. & Wells, J. R. E. *J. Biol. Chem.* **262**, 9664–9668 (1987).
- Porod, G. in *Small Angle X-ray Scattering* (eds Glatter, O. & Kratky, O.) 17–33 (Academic, London, 1982).
- Ibel, K. & Stuhmann, H. *J. molec. Biol.* **93**, 255–265 (1975).

ACKNOWLEDGEMENTS. Dedicated to the memory of Julian Wells, whose gift of clones of the chicken histone genes made this work possible. This work was supported by grants from the NIH and the Office of Health and Environmental Research of the US Department of Energy.

Crystal structure of catechol O-methyltransferase

Jukka Vidgren*†, L. Anders Svensson†
& Anders Liljas†

* Orion Corporation, Orion-Farmos, PO Box 65,
FIN-02101 Espoo, Finland

† Molecular Biophysics, Chemical Center, University of Lund,
PO Box 124, S-221 00 Lund, Sweden

CATECHOL O-methyltransferase (COMT, EC 2.1.1.6) is important in the central nervous system because it metabolizes catecholamine neurotransmitters such as dopamine. The enzyme catalyses the transfer of the methyl group from S-adenosyl-L-methionine (AdoMet) to one hydroxyl group of catechols^{1–4}. COMT also inactivates catechol-type compounds such as L-DOPA. With selective inhibitors of COMT in combination with L-DOPA, a new principle has been realized in the therapy of Parkinson's disease^{5–9}. Here we solve the atomic structure of COMT to 2.0 Å resolution, which provides new insights into the mechanism of the methyl transfer reaction. The co-enzyme-binding domain is strikingly similar to that of an AdoMet-dependent DNA methylase¹⁰, indicating that all AdoMet methylases may have a common structure.

The rat liver COMT gene was cloned, expressed in *Escherichia coli* and the protein purified to homogeneity^{11–13}. The enzyme has 221 amino acids and a relative molecular mass of 24,700 (24.7K). Table 1 summarizes the solution of the crystallographic structure and the refinement of COMT in complex with a competitive inhibitor (3,5-dinitrocatechol), Mg²⁺ and the co-enzyme AdoMet at 2.0 Å resolution.

COMT consists of eight α -helices and seven β -strands and has a typical α/β fold (Figs 1 and 2). The helices $\alpha 4$ – $\alpha 8$ and the parallel β -strands $\beta 1$ – $\beta 5$ are very similar to the nucleotide-binding motif (the Rossmann fold) found in alcohol dehydrogenase (ADH) for example¹⁴. The adenosine of AdoMet superimposes well on the adenosine moiety of NAD⁺ in ADH. The last residue of $\beta 1$ in COMT is Gly 66, which corresponds to a conserved glycine residue in all dehydrogenases (Gly 199 in ADH)¹⁴. The last residue in the $\beta 2$ -strand is Glu 90, which binds to the ribose hydroxyls of AdoMet. In dehydrogenases this residue corresponds to a conserved acidic residue (Asp 223 in ADH). The C-terminal part of the COMT structure consists of two antiparallel β -strands ($\beta 6$ and $\beta 7$).

The active site of COMT includes the co-enzyme-binding motif and the catalytic site situated in the vicinity of the Mg²⁺ ion. Residues from the amino terminal helices, through the nucleotide-binding fold to the loop region of the C-terminal β -strands ($\beta 6$ and $\beta 7$), build up the active site.

AdoMet binds only when the enzyme is complexed with the Mg²⁺ ion¹⁵, but bound AdoMet has no direct interaction with the magnesium. The adenine has favourable edge-to-face van der Waals contacts with Trp 143 (Fig. 2). His 142 and Met 91 have van der Waals interactions with opposite sides of the adenine. The side-chain oxygen of Ser 119 has a hydrogen bond to N6 and its main-chain oxygen binds to N1 of the purine ring.

TABLE 1 Diffraction data, structure solution and refinement

Diffraction data							
	Native I	Native II	Cd	Hg			
Resolution (Å)	2.9	2.0	3.2	2.9			
Total observations	16,855	158,331	10,954	18,712			
Unique reflections	5,944	16,005	4,231	5,540			
Completeness of data (%)	93	85	88	87			
$R_{\text{sym}}\%$	6.6	8.3	6.6	7.8			
Structure solution statistics							
Derivative	Resolution (Å)	Unique reflections	$R_{\text{iso}}(\%)^{\dagger}$	Number of sites	$R_c(\%)^{\ddagger}$	Figure of merit	Phasing power §
Cd	3.2	4231	12.0	1	53	0.64	1.73
Hg	2.9	5540	12.5	1	62	0.60	1.70

Crystallization of COMT has been described¹². The structure was determined using X-ray diffraction data from a Siemens area detector equipped with a 3-circle goniostat mounted on a Rigaku RU200BEH rotating anode. Data reduction was done by the XENGEN program suite²³. The space group is trigonal, $P3_121$, $a = 51.3$ Å, $b = 51.3$ Å, $c = 168.5$ Å. Two heavy-atom derivatives (cadmium and mercury) were obtained by co-crystallization. The CCP4 program suite was used for calculations²⁴. One heavy atom site per derivative was found by difference Patterson techniques. Heavy-atom parameters were refined and phases calculated for both derivatives (overall figure of merit, 0.70). MIR phases resulted in a good electron density map. After solvent flattening with the SQUASH program²⁵, most of the backbone skeleton was interpretable with the 'O' molecular modelling package²⁶ on a Silicon Graphics Personal Iris workstation. This initial model was refined by simulated annealing with all data to 2.9 Å with the program XPLOR²⁷. Native data to a resolution of 2.0 Å were collected on an MAR imaging plate using synchrotron radiation at station 9.5 at the Daresbury Laboratory, UK. Data reduction was done with programs Denzo and Scalepack (written by Z. Otwinowski). The current molecular model of COMT includes 216 of the 221 residues. The last five residues of the C terminus do not have an interpretable electron density; 160 water molecules were included in this model. Individual temperature factors were refined in the final cycle. The crystallographic R -value was 19.4%, including 15,300 reflections between 8 and 2.0 Å resolution. The r.m.s. deviation in bond lengths was 0.013 Å and 2.8° in bond angles. Atomic coordinates will be deposited to the Brookhaven Protein Databank.

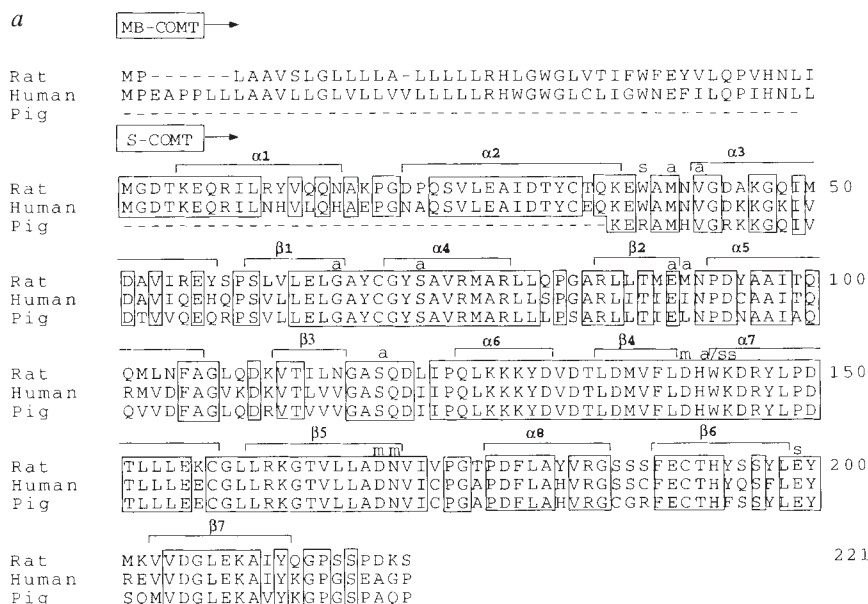
* $R_{\text{sym}}(\%) = 100 \sum_i | \langle I_i \rangle - I_i | / \sum_i I_i$ and $\langle I_i \rangle$ is the average of I_i over all symmetry equivalents.

† $R_{\text{iso}}(\%) = 100 \sum_i | F_{\text{PH}}^2 - F_{\text{P}}^2 | / \sum_i (F_{\text{PH}}^2 + F_{\text{P}}^2)$ where F_{PH} and F_{P} are the derivative and native structure factors, respectively.

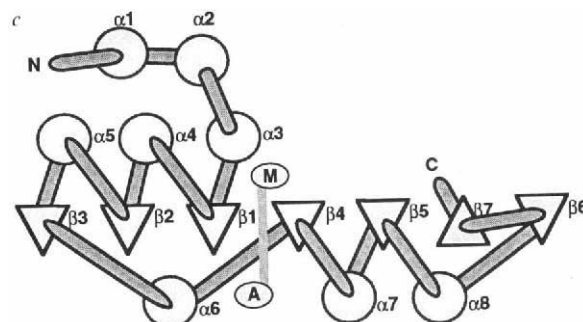
‡ $R_c(\%) = 100 \sum_i | |F_{\text{PH}} \pm F_{\text{P}}| - |F_{\text{H}}(\text{calc})| | / \sum_i |F_{\text{PH}} \pm F_{\text{P}}|$ for centric reflections.

§ Phasing power = $\langle F_{\text{H}} \rangle / E$; the r.m.s. heavy-atom structure factor amplitude divided by the residual lack of closure error.

FIG. 1 a, Comparison of amino-acid sequences of COMT (rat¹¹, human²⁸, pig²⁹). Secondary structure elements in the structure of rat soluble COMT and important active-site residues involved in binding of ligands (a, Adomet; m, magnesium; s, substrate/inhibitor) are indicated. Numbering of residues corresponds to that for the soluble enzyme (S-COMT). Extension of the membrane-bound protein (MB-COMT) consists of the first 50 amino-terminal residues. b, Consensus regions 'a' and 'b' as found by manual alignment of sequences of mammalian methyltransferases²². The corresponding regions of COMT are shown (s, small neutral; o, hydrophobic amino acids, and x, any amino acid). Region 'a' is involved in the binding of AdoMet in COMT. c, Topological diagram of secondary structure elements of COMT. The position of the bound AdoMet is shown (A, adenine; M, methionine).



b	Sequence	COMT sequence number
consensus region a COMT	L D/E o G s G s G L E L G A Y C G	63-70
consensus region b COMT	L R/K P G G x L L R K G T V L	160-166



The ribose ring lies side by side with Trp 143. Met 40 orients the sulphur of AdoMet, with its electron-deficient methyl group, towards the nucleophilic hydroxyl group of the catechol substrate. The carboxyl oxygens of the methionine moiety of AdoMet are hydrogen-bonded to the backbone nitrogen of Val 42 and to a water molecule. The amino group of the methionine moiety has hydrogen bonds to Asp 141, to the main-chain oxygen of Gly 66 and to the side-chain oxygen of Ser 72.

The Mg^{2+} ion bound to COMT has a crucial role in substrate binding and hence in the catalytic activity. As shown in Fig. 3, the Mg^{2+} is coordinated to the side-chain oxygens of residues Asp 141, Asp 169 and Asn 170, and to both inhibitor hydroxyls. The sixth position in the octahedral magnesium coordination is occupied by a water molecule.

The binding site for 3,5-dinitro catechol (a competitive inhibitor of COMT) is in a shallow groove on the surface of the protein. No protruding part of the protein encloses the inhibitor. Both hydroxyl groups of the inhibitor are complexed with the Mg^{2+} ion as shown in Fig. 3. The hydroxyl group in position 1 has an important hydrogen bond to the carboxyl oxygen of Glu 199. The other hydroxyl is near the methyl group donated by AdoMet ($O \cdots C$ distance of 2.6 Å). The 3-nitro group of the inhibitor has favourable van der Waals interactions with Trp 143. Trp 38 is located edge-to-face with the planar structure of 3,5-dinitro catechol, which allows an ideal hydrophobic contact that probably contributes to the specificity for aromatic dihydroxy compounds.

On the basis of the atomic structure of COMT, it can be postulated that the binding of catecholamines is very similar to the binding of the inhibitor. The methyl transfer from AdoMet to the catechol substrate catalysed by COMT is a direct bimolecular transfer of the methyl group from the sulphur of AdoMet to the oxygen of the catechol hydroxyl in an S_N2 -like transition state¹⁶. The exact juxtaposition of the substrate to the methyl group is possible because of the binding of the two hydroxyl groups to the Mg^{2+} ion. The possibility that the methyl transfer reaction is catalysed by a cysteine residue abstracting the proton from the catechol hydroxyl¹⁷ is excluded because there is no cysteine in the active site of COMT. Instead, one hydroxyl of the substrate is surrounded by three positively charged groups inducing it to release its proton to become a negatively charged phenolate ion. These moieties are the Mg^{2+} , the methyl group of AdoMet, and Lys 144. The Mg^{2+} ion in particular probably lowers the pK_a of the hydroxyl group significantly. In contrast, the proton of the other hydroxyl is stabilized by the negatively charged carboxyl group of Glu 199. The ionized hydroxyl makes a direct nucleophilic attack on the electron-deficient methyl of AdoMet. The methyl ether of the catechol substrate subsequently dissociates. Modelling of the *O*-methylated dopamine in the active site of COMT shows that Lys 144 will repel the methyl group of the *O*-methyl ether. COMT is a very slow enzyme, having a

turnover number of 24 per min¹⁸. The binding affinity of the demethylated AdoMet, *S*-adenosyl-L-homocysteine (AdoHcy), is almost the same as that of AdoMet¹⁷. A conformational change may be involved in the dissociation of AdoHcy.

The methylation of catecholamines can occur either at the *m*- or the *p*-hydroxyl group⁴ (corresponding to positions 1 and 2 in the inhibitor structure). The ratio of the *m*-/*p*-methylation depends on the substituents of the catechol ring, the *m*-position being preferred. It has been suggested that the methylation site is determined by the pK_a of the hydroxyl and by the possible stereo-electronic interactions of the substrate side chain with hydrophobic protein residues^{19,20}. From the structure of COMT, it can be concluded that the binding of catecholamines, such

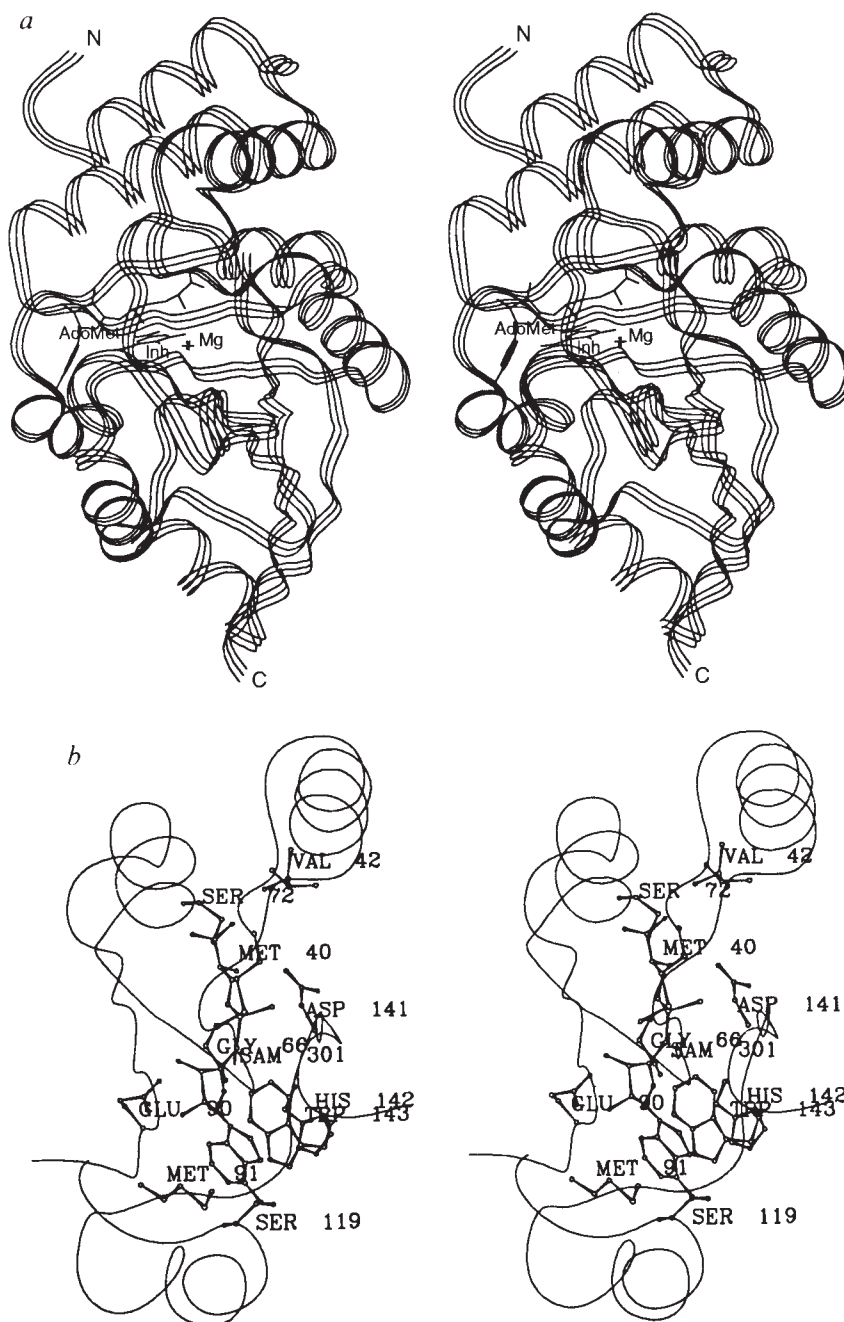


FIG. 2 *a*, Overall schematic stereo view of COMT. The ligands AdoMet, Mg^{2+} and inhibitor (Inh) are shown. *b*, Binding of AdoMet (SAM) in COMT (in stereo). The electron density of the co-enzyme AdoMet was clearly observed in the active site of COMT as a result of co-crystallization. All of the residues interacting with AdoMet and significant parts of the main chain are shown.

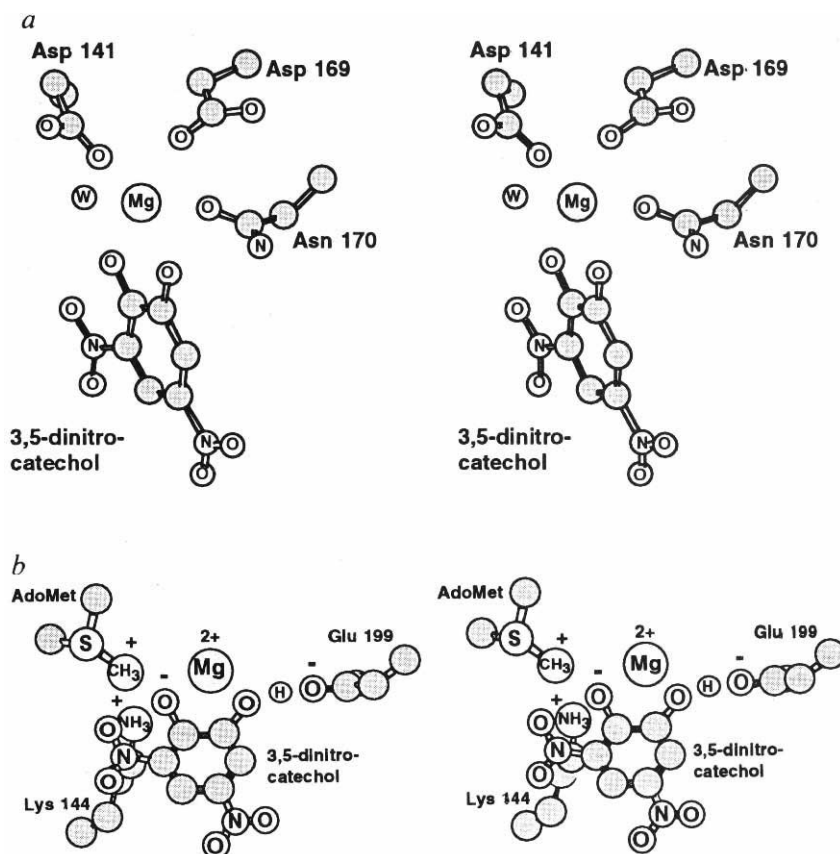


FIG. 3 Stereo view of the active site of COMT. The enzyme was crystallized in complex with Mg^{2+} and the inhibitor 3,5-dinitro catechol, which were clearly defined in the electron density maps. *a*, The binding site for Mg^{2+} . Here the substrate position is occupied by 3,5-dinitro catechol; the sphere marked 'W' represents a water molecule. *b*, Binding of 3,5-dinitro catechol in the active site of COMT. The most important residues that interact with the bound inhibitor are shown.

as dopamine, with the *p*-hydroxyl group toward the methyl of AdoMet may cause a stronger repulsion between the positively charged amino group of the substrate side chain and the hydrophobic residues Trp 38 and Pro 174.

The amino-acid sequence of soluble rat COMT is 81% identical with the human enzyme. The alignment of the known COMT sequences can be seen in Fig. 1. Almost all active site residues identified for rat COMT are conserved. The most interesting difference is Arg 38 in the pig protein (Trp in human and rat COMT). The pig enzyme binds the substrate and nitro catechol inhibitors with an affinity ten to one hundred times less than the rat enzyme²¹. Evidently Trp 38 is important for high-affinity binding of the catechol compounds.

As a result of manual alignment of primary structures, two consensus sequences of methyltransferases have been found²² (Fig. 1). Consensus region 'a'²² includes Gly 66 in the turn between β 1 and α 4 at the binding site for AdoMet. This is generally the location for consensus residues in nucleotide-binding proteins. Residues corresponding to Glu 90 are negatively charged, at least in some of the AdoMet-utilizing enzymes. The location of the consensus element in small molecule methylases is about 60 residues from the N terminus, suggesting that not only is the AdoMet-binding fold common to these enzymes, but that the three helices preceding β 1 are as well.

The crystal structure of a DNA methylase has recently been described¹⁰. The AdoMet-binding motif (the Rossmann fold) of this enzyme and that of COMT are similar. The loop region (corresponding to β 1a4 in COMT) containing the important glycine residue is similar in the DNA methylase. The purine and ribose rings in both structures have hydrophobic contacts to a tryptophan residue. There is a significant difference between these enzymes in the binding of the ribose moiety of AdoMet. The important hydrogen bonds between Glu 90 and the ribose hydroxyls in COMT are realized by the hydrogen bond between the ribose hydroxyl and the main-chain nitrogen in the DNA

methylase. The substrate-binding parts of these methyltransferases are naturally different. The two similar structures and the consensus elements of AdoMet-utilizing enzymes may be as significant as they are in other nucleotide-binding proteins and indicative of a similar structure and divergent evolution for all AdoMet-dependent methylases. □

Received 25 October 1993; accepted 27 January 1994.

1. Axelrod, J. & Tomchick, R. *J. biol. Chem.* **233**, 702–705 (1958).
2. Ball, P., Knuppen, R., Haupt, M. & Breuer, H. *J. clin. Endocr.* **34**, 736–746 (1972).
3. Borchardt, R. in *Enzymatic Basis of Detoxification* (ed. Jakoby, W. B.) 43–62 (Academic, New York, 1980).
4. Guldberg, H. & Marsden, C. *Pharmac. Rev.* **27**, 135–206 (1975).
5. Bäckström, R. *et al. J. med. Chem.* **32**, 841–846 (1989).
6. Borgulya, J., Bruderer, H., Bernauer, K., Zurcher, G. & Da Prada, M. *Helv. chim. Acta* **72**, 952–968 (1989).
7. Männistö, P. & Kaakkola, S. *Trends pharmac. Sci.* **10**, 54–56 (1989).
8. Linden, I.-B. *et al. Pharmac. exp. Ther.* **247**, 289–293 (1988).
9. Männistö, P. *et al. in Progress in Drug Research* (ed. Jucker, E.) 291–350 (Birkhäuser, Basel, 1992).
10. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. & Roberts, R. *Cell* **74**, 299–307 (1993).
11. Salminen, M. *et al. Gene* **93**, 241–247 (1990).
12. Vidgren, J., Tilgmann, C., Lundström, K. & Liljas, A. *Proteins* **11**, 233–236 (1991).
13. Lundström, K., Tilgmann, C., Peränen, J., Kalkkinen, N. & Ullmanen, I. *Biochim. biophys. Acta* **1129**, 149–154 (1992).
14. Eklund, H. *et al. J. molec. Biol.* **146**, 561–587 (1981).
15. Jeffery, D. & Roth, J. *Biochemistry* **28**, 2955–2958 (1987).
16. Woodward, R., Tsai, M., Floss, H., Crooks, P. & Coward, J. *J. biol. Chem.* **255**, 9124–9127 (1980).
17. Coward, J., Slisz, E. & Wu, F. *Biochemistry* **12**, 2291–2297 (1973).
18. Schultz, E. & Nissinen, E. *Biochem. Pharmac.* **38**, 3953–3956 (1989).
19. Creveling, C., Morris, N., Shimizu, H., Ong, H. & Daly, J. *Molec. Pharmac.* **8**, 398–409 (1972).
20. Thakker, D., Boehlert, C., Kirk, K., Antkowiak, R. & Creveling, C. *J. biol. Chem.* **261**, 178–184 (1986).
21. Piedraffita, J., Elorriaga, C., Fernandez-Alvarez, E. & Nieto, O. *J. Enzym. Inhib.* **4**, 43–50 (1990).
22. Gomi, T., Tanihara, K., Date, T. & Fujioka, M. *Int. J. Biochem.* **24**, 1639–1649 (1992).
23. Howard, A. *et al. J. appl. Crystallogr.* **20**, 383–387 (1987).
24. CCP4, A Suite of Programs for Protein Crystallography (SERC (UK) Collaborative Computing Project No. 4, Daresbury Laboratory, Warrington, 1979).
25. Cowtan, K. & Main, P. *Acta crystallogr. D* **49**, 148–157 (1993).
26. Jones, T. A., Zou, J.-Y., Cowan, S. & Kjeldgaard, M. *Acta crystallogr. A* **47**, 110–119 (1991).

27. Brunger, A. X-PLOR Version 3.0, A System for Crystallography and NMR (Yale University, CT, 1992).
28. Lundström, K., Salminen, M., Jalanko, A., Savolainen, R. & Ulmanen, I. *DNA Cell Biol.* **10**, 181–189 (1991).
29. Bertocci, B. et al. *Biochim. biophys. Acta* **1080**, 103–109 (1991).

ACKNOWLEDGEMENTS. This work was supported by the Nordic Industrial Fund (J.V.), Nordic Molecular Biology Association (NOMBA) (J.V.) and Swedish Natural Research Council (NFR) (A.L.). Equipment was obtained with funds from the Swedish Council for Planning and Coordination of Research, the SE bank, and the Knut and Alice Wallenberg foundation. We thank the Daresbury Laboratory and S. M. McSweeney for synchrotron radiation facilities and for help with data reduction.

ERRATUM

Parasitism, mutation accumulation and the maintenance of sex

R. Stephen Howard & Curtis M. Lively

Nature **367**, 554–557 (1994)

DURING the production process, the contrast in shading in Figs 1 and 2 of this letter was lost; in addition, parts *a* and *b* of Fig. 3 were transposed. These figures should have appeared as they are shown here. The first sentence of each legend is given for guidance. □

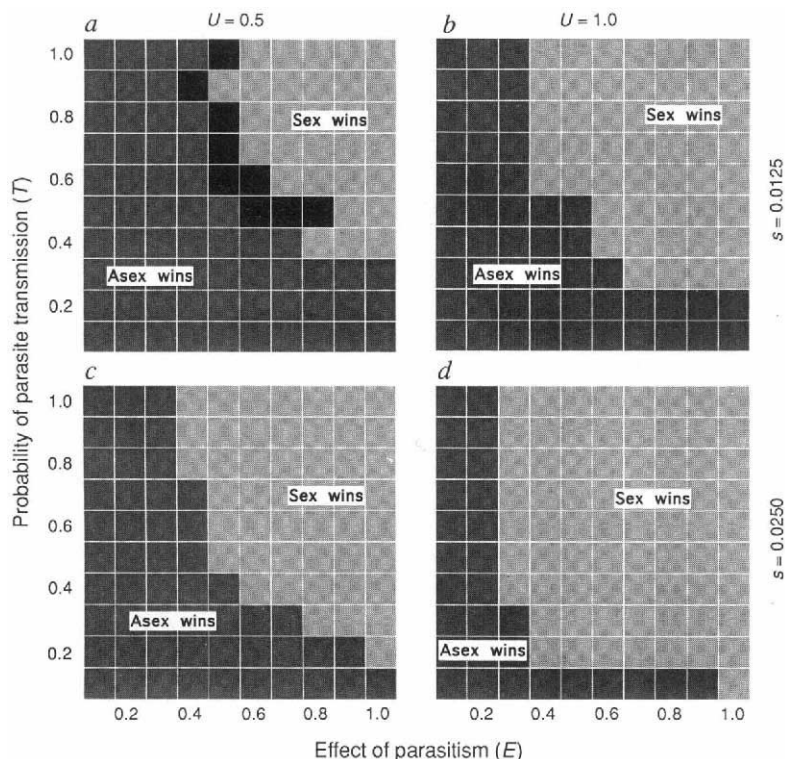


FIG. 2 Results from computer simulations in which sexual populations were challenged by monotypic asexual lineages in the presence of coevolving parasites and selection against harmful recurrent mutations.

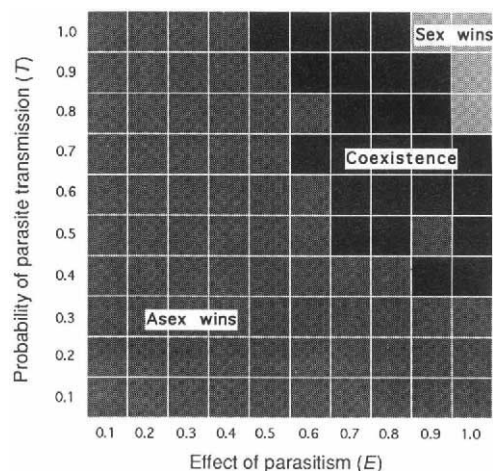


FIG. 1 Results from computer simulations in which sexual populations were challenged by asexual lineages in the presence of coevolving parasites.

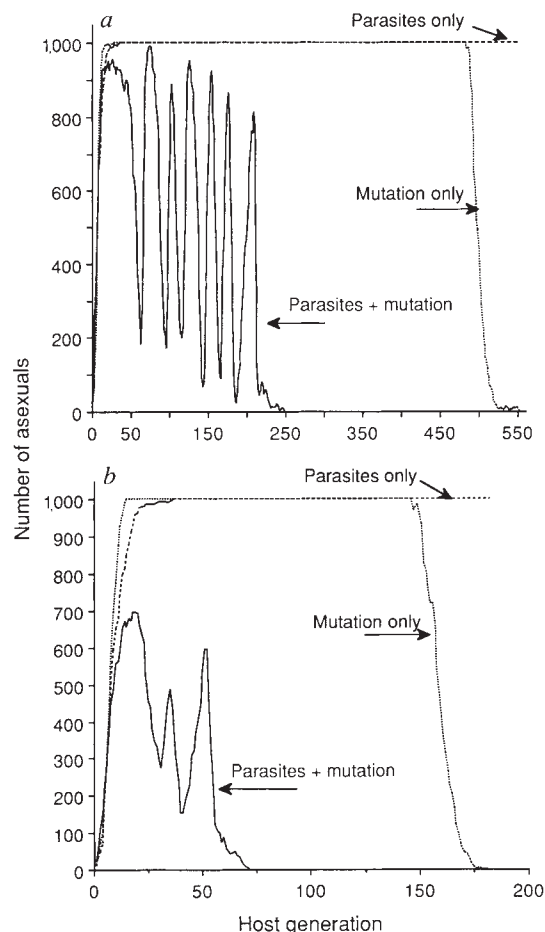


FIG. 3 Population dynamics for monotypic asexual lineages in competition with sexual populations in the presence of either parasites or mutation, or both.

DNA in the mind's eye

Richard L. Gregory

The Astonishing Hypothesis: The Scientific Search for the Soul. By Francis Crick. Scribner's: 1994. Pp. 317. \$25. To be published in the United Kingdom by Simon and Schuster on 19 May at £16.99.

FROM the extraordinary triumph 40 years ago of unravelling the structure of DNA, Francis Crick turned to questions of how the brain works. It is surely no coincidence that over the past few years the institute to which he moved, the Salk Institute in La Jolla, has with the University of California at San Diego become an important magnet for brain research, attracting established and up-and-coming scientists of all kinds — and brain research requires all kinds — as well as philosophers prepared to learn anatomy and physiology and to consider the results of experiments, including psychological phenomena.

Here Crick sits at the centre of a vibrant web. When he tweaks a thread, people will travel any distance to hear and be heard by him, sometimes only to be told to think again. These Sun-worshippers are prepared to ask difficult questions and to seek answers wherever they may be found, whether in the details of molecular chemistry or in the structure and function of the cunning cortex.

A shared vision of these travellers is that of understanding vision. The way in which we see was contemplated by the Greeks, but only since the middle of the nineteenth century has the scientific method been applied consistently to the problem, and there is still a lack of agreement on how best to discover the processes underlying vision and consciousness. This book is an account of this dramatic enterprise by one of the greatest living biologists, written with clarity and charm.

There is, to be perhaps too frank, a general feeling that Nobel laureates who change tack in mid-life are loose cannons. But Crick has chosen his targets with considerable care, and has succeeded in piercing many revealing holes through barriers of puzzlement and ignorance. He sets his sights firmly on neurons, believing that a great deal of what other people consider to be important about the brain and consciousness is irrelevant, misleading or a waste of time. Quantum mechanics, for example, is judged to be neither here nor there: although quantum principles are necessary for an understanding of

how snooker balls maintain their shape and bounce about, the basic principles of matter can and should be ignored. For Crick, the actors in the play of the brain are the billions of interconnected neurons and their electrochemical signals. (Perhaps the actors are rule-based symbols, although Crick doesn't move far in the direction of symbolic models of cognition, no doubt wisely



for his own purposes.)

The "Astonishing Hypothesis" is that there is no homunculus, or little person, inside our heads, constituting the self; that there is no extra-vehicular soul, only the immensely complex and currently little understood interactive activity of neurons. Thus, as Julian Offray de la Mettrie said in 1748 (and he was thrown out of his medical practice in Paris for saying it), man is a machine. As we come to see this, the limitations of what is meant by 'machine' become relaxed, but not so far that anything goes, or rather that anything is accepted. What is lost is Descartes' notion of an immaterial mind that lies beyond scientific explanation. Yet even this position has its semantic problems, for what might the term 'scientific' accommodate in a hundred years from now? This uncertainty is why Crick calls

his view a hypothesis: he concedes that his neuronal account remains unproved and admits that he lacks a theory of consciousness — of why red is red.

We are guided through the delightful intricacies of the psychology of vision (especially the filling-in of blind regions and phenomena of ambiguity where perception takes off from stimuli); attention and memory; the structure of the brain; and the workings of its neurons. There is a useful chapter on the power and weakness of experimental techniques, including microelectrode recording, patch-clamping and the latest imaging methods. Effects of brain damage are also discussed. All this leads to potentially important speculations, the upshot being that consciousness remains mysterious. But Crick will continue to contemplate the future of brain science — the detailed and comprehensive study of neurons.

What he seeks are neural correlates of sensations, thoughts and perceptions: that is, brain activity likely to be related to 'qualia', the subjective quality of mental experience. If these associations are found, a philosopher might still say: "So what?" Crick has no time for philosophers, at least those without knowledge of neuroanatomy and function (so he values Patricia and Paul Churchland's contributions), but even he would probably have to admit that the age-old question of the nature of consciousness remains, for correlations can never directly and unambiguously provide answers or prove cause and effect. Day is followed by night, but it is nonsense to say that one is caused by the other — a proper understanding of the phenomenon involves a

leap beyond everyday experience to a mental picture of the Earth spinning around the Sun. By analogy, some philosophers would say that Crick is too 'Earth-bound' to explain how qualia are related to (caused by?) neuronal activity.

Crick dedicates his book to Christof Koch, who has worked with him in developing many of the ideas it contains. They both agree with the psychologist V. S. Ramachandran that:

visual perception does not involve intelligent deduction of the type we use in constructing an argument, nor does it involve the vague idea that the brain simply 'resonates' to the visual input. Neither does it require the solving of elaborate equations, as often implied by AI [artificial intelligence] researchers. Instead [Ramachandran] believes perception "uses rules of thumb, short-cuts, and clever sleight-of-hand tricks that are acquired by trial and

error through millions of years of natural selection. This is a familiar strategy of biology" [p. 77].

The digital computer is rejected as a model of the brain in favour of a scheme involving many special processors probably consisting of interacting nets of neurons. Cortical organization is seen as being dictated by the need for short pathways and the efficient use of a limited (though enormous) number of neurons. This wiring requirement produces specialized, highly interactive modules linked to nearby functionally related ones, like 'villages' within a city.

Crick argues that far more detailed knowledge of human brain anatomy — particularly the cortical layers — is required to make any real progress in locating the main sites of possible "awareness neurons". His favourite candidates are the inner cortical layers and subcortical structures, especially the 'gateway to the cortex', the thalamus. And because many specialized visual processors deal separately with form, movement, depth, colour and so on, there is the 'binding problem' of how their contributions come together to give coherent, vivid perceptions. Crick thinks that neural mechanisms of attention may hold the key, with neural oscillations as the basis of selective attention and captured synchronous outputs of processors underlying binding.

But what of qualia? Why is red red? Crick would seem to agree with Stuart Sutherland's statement: "Consciousness is a fascinating but elusive phenomenon; it is impossible to specify what it does, or why it evolved. Nothing worth reading has been written on it." But by seeking neural correlates, he hopes to capture consciousness from philosophy. And by rejecting special properties (as in vitalism), he emphasizes emergence, pointing out that novelty generated by combinations of particles, or properties, even at the molecular level, is the essence of chemistry and life. Here 'emergence' does not mean mysteries popping out of the undergrowth; it means that with a sufficient understanding of interactive processes, we should come to understand why a complex whole has properties its parts lack on their own, and how the parts are modified by the context in which they lie.

This is a flexible kind of atomism. Perhaps the ultimate aim is to derive a conceptual model for understanding the emergence of consciousness from the physics of neurons. All this from DNA, through the evolution of intelligence. Dare we call it Darwinian Neural Awareness? □

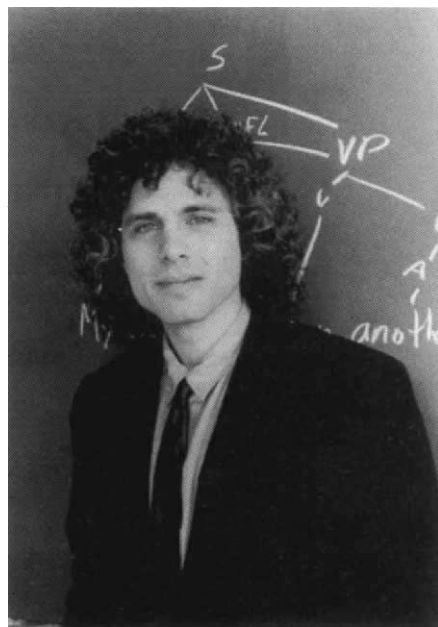
Richard L. Gregory is in the Department of Psychology, University of Bristol, Bristol BS8 2TH, UK.

The talking ape

Christopher Longuet-Higgins

The Language Instinct: How the Mind Creates Language. By Steven Pinker. *Morrow: 1994. Pp. 494. \$23. To be published in the United Kingdom by Viking on 11 April at £20.*

HERE at last is a marvellously readable book about language, written by a real expert. Steven Pinker tackles with wit and erudition the kinds of question everyone asks about language. Why do immigrants have such trouble learning a new tongue, only to have their children ridicule their grammatical mistakes? Why is the Maple Leafs hockey team not called the Maple Leaves? Have scientists actually succeeded in reconstructing the first language



Pinker — no linguistic hang-ups.

spoken on Earth? Are there genes for grammar? Can chimpanzees really learn human sign languages? How could language have evolved? Does a person's language control the way he or she thinks? Is the English language gradually falling apart?

To these questions Pinker brings not only an expertise in linguistics and psychology and a wide knowledge of biology, but also an ability to understand the ordinary person's linguistic hang-ups and to shake them loose with gentle ridicule. Here (slightly condensed) is the opening of his chapter on "The Language Mavens" (the word 'maven' was coined by the critic William Safire for a self-appointed regulator of the English language):

Imagine you are watching a nature documentary. The video shows the usual gorgeous footage of animals in their natural habitats. But the voiceover reports some

disturbing facts. White-crowned sparrows carelessly debase their calls, and the song of the humpback whale contains several well-known errors. Your reaction would probably be, 'What on earth could it mean for the song of the humpback whale to contain an "error"? Who is this announcer, anyway?

All the best writers in English, at all periods, including Shakespeare and most of the mavens themselves, have been among the most flagrant flouters [of the mavens' rules]. The rules conform neither to logic or to tradition, and if they were ever followed they would force writers into fuzzy, clumsy, wordy, ambiguous, incomprehensible prose, in which certain thoughts are not expressible at all.

The principal theme of *The Language Instinct* is implicit in its title. Human language defies explanation as the automatic response of a general-purpose brain to the complex stimuli of early childhood; it must, as Noam Chomsky has always insisted, be a largely inherited faculty. But while Chomsky is fiercely agnostic about the evolutionary origin of language, Pinker makes a convincing case that language evolved by a fully Darwinian mechanism, like the elephant's trunk. (Those who have lost their childhood wonder at this astonishing proboscis will regain it as soon as they read the breathtaking first paragraph on page 332.)

There are other points at which Pinker departs from Chomskian orthodoxy, most notably in postulating a mental language — "mentalese" — in which we marshal our thoughts before expressing them in, say, English, and into which we translate what other people say to us. The value of such a hypothesis depends, of course, on its predictive power, and one may prefer an account of thinking more along the lines laid down in Philip Johnson-Laird's influential book *Mental Models* (Harvard University Press/Cambridge University Press, 1983). But at least the concept of an internal language of the mind seems preferable to the simplistic idea that Englishmen think in English while Frenchmen think in French — a hypothesis that creates acute problems about how chess players or tennis players or flute players might think, and how people manage to translate, however awkwardly, from one language into another.

Whatever its eventual impact on linguistics and psychology, *The Language Instinct* will undoubtedly be greeted as a distinguished contribution to the lay understanding of science. In the past, linguists have been less effective as communicators than, say, astronomers or biologists, but now the educated layman will have no excuse for imagining that 'deep structure' is essentially the same thing as 'universal grammar'. (There is an excellent glossary in which such misconceptions can be quietly sorted out.) And Pinker is determined to give us some hard, detailed information to get our minds

into. In the Bantu language Kivunjo, he tells us, the verb form *Näikimlyiä*, meaning "He is eating it for her", is one of about half a million possible forms, which native speakers have no trouble in assembling on the fly from bits and pieces such as these:

- N- A marker indicating that the word is the 'focus' of that point in the conversation.
- ä- A subject agreement marker, identifying the eater as falling into class 1 of the 16 gender classes, 'human singular'. Other genders (kinds, not sexes) embrace nouns that pertain to several humans, thin or extended objects, objects that come in pairs or clusters, the pairs or clusters themselves, instruments, animals, body parts, diminutives (small or cute versions of things), abstract qualities, precise locations and general localities.
- i- Present tense. Other tenses in Bantu can refer to today, earlier today, yesterday, no earlier than yesterday, yesterday or earlier, in the remote past, habitually, ongoing, consecutively, in the future, at an indeterminate time, not yet and sometimes.
- ki- An object agreement marker, in this case indicating that the thing eaten falls into gender class 7.

And so on.

One of the mysteries Pinker does not resolve for the inquisitive reader is the precise relationship between the early- and the late-flowering varieties of Chomskian linguistic theory. His reasons may be diplomatic, or simply expository: a fear of the pedagogic confusion that might result from attempting to superimpose the most recent version of "GB theory" on the "Xbar theory" to which he has so carefully introduced the reader. Perhaps he merely felt it wiser to hide the antics of the professional linguists from the searching gaze of student innocence.

For all other readers, *The Language Instinct* will illuminate every facet of human language: its biological origin, its uniqueness to humanity, its acquisition by children, its grammatical structure, the production and perception of speech, the pathology of language disorders and the unstoppable evolution of languages and dialects. With its wealth of examples, its flawless typesetting, its wide-ranging bibliography and its irresistible good humour, Pinker's book is certain to increase its readers' respect for the amazing natural phenomenon that the author and his colleagues have made their life's (to a maven, lives'?) study. □

Christopher Longuet-Higgins is in the Laboratory of Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Reactions speak louder than words

Roald Hoffmann

La Parole des Choses. By Pierre Laszlo. Hermann: 1993. Pp. 320. FF160.

THERE has always been a link between language and chemistry. Lavoisier begins his revolutionary *Traité Élémentaire de Chimie* with a quote from the Abbé de Condillac: "We think only through the medium of words. Languages are true analytical methods." Lavoisier then reflects on his own work: "Thus, while I thought of myself employed only in forming a Nomenclature... my work transformed itself by degrees, without my being able to prevent it, into a treatise upon the Elements of Chemistry". The distinguished European writer Elias Canetti, author of a remarkable study of mass behaviour, *Crowds and Power*, and a striking novel of the 1930s, *Auto da Fé*, earned a PhD in chemistry. He credits chemistry with teaching him the importance of linguistic structure. And Benjamin Lee Whorf, the great American linguist who made a case for language shaping culture, trained as a chemical engineer at the Massachusetts Institute of Technology. Whorf was not averse to "an occasional chemical simile". In an essay on languages and logic he writes: "the way the constituents are put together in these sentences of Shawnee and Nootka suggests a chemical compound, whereas their combination in English is more like a mechanical mixture".

Pierre Laszlo's rich and original book explains the link between language and chemistry. Laszlo makes an analogy between molecules and their transformations on one hand and linguistic structures such as morphemes, phonemes, ideograms and pictograms, transformations of mode and description on the other. If the latter sound abstruse to a chemist, imagine what sense a functional group, elementary reactions, C_6H_6 or a benzene ring drawn out (of course, without a C or H indicated) would make to a linguist other than the late Whorf.

La Parole des Choses (perhaps one way to translate the title is "The Way Things Speak", but that doesn't capture the allusion to the structuralist and literary theory use of *parole*) is actually two books. The working out of the language-chemistry analogy is one theme that runs through this readable volume. Another theme, interwoven, is simply that of the beauty of chemistry. The logic of the molecular science is masterfully explicated, with few compromises, for an intelligent reader outside chemistry. For instance, the book contains a beautiful discussion of nitrogen inversion, the best short introduction to nuclear magnetic resonance I've read and

even a brief account of femtosecond reactions. But consistent with Laszlo's broad cultural theme, these expositions are juxtaposed with, respectively, a quote from Gérard de Nerval (a nineteenth-century French poet and writer), a touching retelling of some of Laszlo's misadventures in taking his first NMR spectrum in 1961, and an excerpt from Blake's *Milton*.

Insights and information abound. The linguistic analogy builds convincingly, but one does not have to buy into it to appreciate the capsule description of the economics of the perfume industry (and the composition of "Obsession") or the surprising use by Hegel of Berzelius's electrochemical theory or the prolonged alchemical and chemical use of the word *menstruum* for solvent.

Let me see if I can put my finger more precisely on the value in this book. To deconstruct scientific texts is so easy, too easy. Ditto for the use of literary allusions to pretty up science writing. Laszlo does something more — he sets up an intriguing analogy, of the *processes* of language and chemistry, and lets the analogy illuminate each field. □

Roald Hoffmann is in the Department of Chemistry, Cornell University, Ithaca, New York 14853-1301, USA.

Linguistic geography

Mark Pagel

Atlas of the World's Languages. Edited by Christopher Moseley and R. E. Asher. Routledge: 1994. Pp. 372. £395, \$599.95.

FIFTEEN years ago, one respected encyclopaedia of the world's languages placed the number of known extant human languages at around 4,550. By 1984 a new survey found that at least 5,450 different languages were being spoken on Earth. The number rose to 6,170 by the time of the appearance of the eleventh edition of *Ethnologue: Languages of the World* (edited by B. F. Grimes, Summer Institute of Linguistics, 1988), and to more than 6,500 by the twelfth edition. The true number is probably even higher. This diversity of languages, all of them spoken by the same species, is testimony to the human — one might say — *compulsion* to speak. Unlike chimpanzees, which need daily operant prodding to use even a small repertoire of words, it is difficult to stop humans from using language: linguists

have even observed groups of young children spontaneously create fully fledged languages. Whoever said that our ability to use language is not an adaptation?

The earliest language was probably first spoken sometime within the past 100,000 years. Nobody knows what this language was like, but it may have left traces: a few sounds seem to recur in most languages. The most celebrated is 'tik' which seems to be associated with words for 'one', 'finger' or 'digit' in many languages (toe?). Nevertheless, languages are in a constant state of Heraclitean flux and they evolve at a much higher rate than most biological traits. A rule of thumb is that a language will replace 15–20 per cent of its fundamental vocabulary per millennium with



new or non-cognate (unrelated) words. Two 'sister languages' that stem from a common ancestor acquire differences between them at twice the rate, and may become mutually unintelligible in about 500 years. Subtle differences may appear rather sooner. The English do not mind being asked what time they would like to be 'knocked-up' in the morning, but an American will blush. A conversation with the Venerable Bede would be even more challenging.

The fast pace of language evolution probably means that hundreds of thousands of different languages have been heard on Earth since humans started talking. Most of these will have either evolved into something new over time or have been replaced by a different language. But the rate at which languages become extinct may be dramatically increasing. Some linguists believe that as many as 3,000 languages will become extinct over the next century. The causes are obvious. A few languages, some owing to colonialism and trade, others to despotism, yet others to human biological fertility, have become ascendant: three-quarters of the world's population speak a language from a list of 20. Where English is spoken, typically between 80 and 90 per cent of the native languages have been lost. In Russia, some 70 per cent of the indigenous languages are moribund.

Routledge's monolithic new *Atlas of the World's Languages* provides an unprecedented account of the linguistic composi-

tion of the world. Many previous books on the world's languages have adopted the title of encyclopaedia, but this atlas is the first to provide maps showing the geographical reach of each language. Each of the 135 colour maps contains a numbered and colour-coded key. Those for the Americas and Australia record the languages as they probably were at the beginning of European colonial expansion, and as they are now. These 'time-of-contact' maps will allow scholars to investigate the way in which language groups contract, or in some cases adapt, in response to an incoming language and society. Accompanying the maps are accounts of the linguistic history of each area, probable relationships among the languages and estimates of the number of contemporary speakers based on recent surveys. Not surprisingly, a work of this magnitude is the result of the efforts of many people: Moseley and Asher assembled 14 other linguists, along with numerous collaborators, to contribute to the volume during a gestation of several years.

So large a contribution is the atlas that I predict it will stimulate its own cottage industry of geographical linguistic research. The world's Tower of Babel is the island of New Guinea with more than 1,000 different languages spoken in 310,000 square miles. Compare this to China with only 90 languages in an area 12 times the size. One of those 90, Mandarin, is spoken by 711 million people. Africa contributes about 2,000 languages, the Americas another 900, whereas Europe and the Middle East have only about 300. Maps of Australia show that more than 400 aboriginal languages were spoken at the time of European contact. Now, by my counts from the atlas's tables of numbers of speakers, at least 205 are either extinct or have so few speakers that the number is unknown. A further 150 are spoken by fewer than 30 people, an amount below which linguists consider a language not to be viable. Dividing Australia into northern and southern regions using the atlas's maps makes clear that the loss of linguistic diversity is much more pronounced in the south — the areas of greatest European settlement in Australia. In Alaska, only two language subgroups — Central Yupik Eskimo and Siberian Yupik Eskimo — are still taught to children. Another 45 or 50 native Alaskan or northern Siberian languages are either extinct or may become so in a generation.

Languages and language evolution do not depend for their own survival on the survival of their biological carriers. Like viruses infecting new bodies, elements of language can quickly jump from mind to mind, changing along the way. This is at least partly why languages evolve so quickly. But this also accounts for why a language can equally quickly displace another, even if there are no changes to

the population of speakers. The loss of any human language is the loss of a communication system as fully developed as one's own — all the different human languages are regarded as roughly equally complex. E. Sapir and B. L. Whorf advanced the idea that language structures the mind. One's language is not only the language in which one thinks, it influences the way in which one thinks. Sapir said that "we see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation". An English-speaking person's mind is different from that of a French speaker, and that of a German speaker, and that of the few remaining speakers of North Frisian, a phenomenon that must contribute a certain *ennui* to European Union negotiations. Sapir and Whorf were criticized for their idea earlier this century, but it has now caught the imagination of linguists and philosophers of mind.

So the loss of a language is the loss of a 'way of mind' for its speakers. The last speaker of a language (and there are many) must confront a stark and wretched isolation. The loss of a language is also the loss of a dimension of consciousness: there are 80 consonants in Ubykh and the Pintupi have at least ten words for a hole in the ground. Some linguists believe that as few as 500 languages may survive the twenty-first century. In not too many years it may be apt to describe the world as narrow-minded. □

Mark Pagel will shortly join the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Diversity of life

Michael Taylor

The Common But Less Frequent Loon and Other Essays. By Keith Stewart Thomson. Yale University Press: 1994. Pp. 186. \$25, £17.95.

To some Scots (as once to Shakespeare) a loon is the young farmworker whose pre-machine lifestyle was recorded in David Cameron's *The Ballad and the Plough*. Not inappropriately, either: in this collection of essays, drawn in part from Thomson's "Marginalia" column in *American Scientist*, Thomson turns his hand from electric fish to mass extinctions to nineteenth-century zoology just as the loon (or loun) went from hay-making to swede-singling, ploughing to harvesting.

Actually, Thomson's loon inhabits not Scottish 'fermtouns' but North American lakes. Thomson, a native Briton, is now president of the Philadelphia Academy of Natural Sciences, so of course the 'common loon' is *Gavia immer*, a large and

attractive waterbird, better known to the British as the great northern diver (and hero of Arthur Ransome's children's book *Great Northern*). Thomson discusses the role of the loon, his backyard equivalent of the giant panda or tiger, as a flagship species, a marker for environmental disturbance (which is why this once Common species is Less Frequent).

Indeed, are we doing enough to fulfil our nineteenth-century predecessors' perception of the need to collect, record and then understand the diversity of nature, especially now that the rate of anthropogenic extinction is so much higher? It will do the average physicist good to see how biological and palaeontological research, too often condemned as 'stamp collecting' (itself, as Thomson notes, only too common in the physical sciences), answers and raises further profound and difficult questions. Meanwhile, Thomson's concern about the uncertain future of palaeontology (and, by extension, biological systematics), and more particularly the infrastructure of collections and taxonomic teaching and research, has increasingly been taken up into serious public debate.

Thomson himself is far from being elderly, as anyone who has met him knows. Yet it is a shock to realize that he crossed the Atlantic not in a Boeing 747 but in the *Mauretania* (a pre-1914 Cunard steam turbine liner now little more than some wood panelling in a Bristol wine bar). His autobiographical sketch of how a boy became a scientist restates the need for personal inspiration and encouragement from real teachers.

Thomson is so succinct that one is sometimes left wanting more, although this is a useful reminder of the merits of brevity. Whether one prefers his mannered essay-like style, supplemented by Linda Thomson's tenebrous drawings, to Dawkins or Desowitz, Gould or Thomas, is a matter of taste, but it all adds to authorial biodiversity. One quibble: non-scientists will enjoy most of the book, but not quite all — a nonbiologist may be defeated by the unillustrated discussion of the embryonic neural crest, that surprising marker of a true vertebrate.

Meanwhile, Thomson takes time off from the problems of science and the biosphere to take a simple pleasure in diversity. I doubt whether anyone is familiar with every topic here: the lost tree *Franklinia*, the mysterious fossil *Palaeospondylus* (fish? amphibian?) found in great abundance in just one Scottish quarry, Gilbert White, Piltdown Man, horse riding and current thinking on tetrapod origins, to name a few. Try these essays for bedtime: you can't go far wrong. □

Michael Taylor is in the Department of Geology, National Museums of Scotland, Chambers Street, Edinburgh EH1 1JF, UK.

Beastly beliefs

Mary Midgley

Animal Minds and Human Morals: The Origins of the Western Debate. By Richard Sorabji. Duckworth: 1993. Pp. 267. £35, \$39.95. Distributed in the United States by Cornell University Press.

READERS of *Nature* will agree that it is a good thing to discover reasons for our beliefs, even when we are sure about them already. But discovering reasons can sometimes be surprisingly hard and — still more annoyingly — sometimes the reasons already given turn out to be bad ones, so that we must start the job afresh. That was what happened to Richard

can string the signs together syntactically. This is a point of the highest scientific interest but of no moral relevance whatever. (Italics added.)

The quest, in fact, turned out not to be just an historical one but a live issue. It is indeed astonishing how, when one looks at these arguments, one finds the same confusions that plagued the Greeks still present, fossilized and essentially unchanged, at the core of much present-day thinking. They still distort our ideas, not just about animals, but human psychology and our relationship with nature.

Centrally, the Greeks' tremendous exaltation of the human intellect inclined them to divorce it altogether from the physical world. This exaltation of thought has of course brought us untold benefits, including the invention of science. But it



A witch feeding familiars — "A Rehearsall both straung and true, of hainous and horrible actes committed by Elizabeth Stile, Alias Rockingham, Mother Dutten, Mother Devell, Mother Margaret, Fower notorious Witches, apprehended at winsmore in the Countie of Barks, and at Abbingdon arraigned, condemned, and executed on the 26 daye of Februarie laste Anno. 1579. London". Taken from *Animals and Human Society: Changing Perspectives* ed. A. Manning and J. Serpell (Routledge, £35).

Sorabji when he started to investigate the reasons that Greek philosophers cited for their belief that humans may do anything they like to animals. As he says:

I began my reading with only a historical interest in the large, and largely uncharted, ancient debate on human and animal psychology. But I was led to appreciate that there was a real, live moral problem by the badness of the arguments for a major difference between animals and man. It sounded grand enough when Aristotle and the Stoics declared that man had reason and animals had not. But, as the debate progressed, it began to appear that animals might lack only certain kinds of reasoning, and a stand was taken on their not having speech. When this defence too began to be questioned, a retreat was made to the position that they lacked syntax. 'They lack syntax, so we can eat them,' was meant to be the conclusion. It was amazing to find that modern discussions had reached exactly the same point as the ancient ones. . . . The debate on the ability of chimpanzees to use sign language has come down to the question whether they

distorted the view held of our own species. It moved *Homo sapiens* away from his shaky, midway status between beasts and gods into a closer and much more ambitious alliance with the latter. Human reason, now hailed as divine, increasingly tried to distance itself from its poor relations both in other human faculties and in the outside world. Any mental likeness between humans and animals, which still formed the irremovable lower pole of the contrast, embarrassed this attempt terribly. Likenesses, therefore, were strenuously minimized.

Sorabji shows how this bias led even Aristotle, who as a biologist stressed the continuity between all lifeforms, to make unreal distinctions between the obviously intelligent behaviour of some animals and very similar behaviour in humans. Using the Greek equivalent of modern shudder-quotes, Aristotle kept ruling that such likenesses must somehow be just analogies. Here, already, the concept of reason

was becoming distorted by idealization, and the abstract, unreal notion of a standard, thought-free 'animal' was supplying a simple contrast to it.

The Stoics, however, carried this process much further, and the Stoic view, which was strongly biased by moral considerations, was the one that St Augustine incorporated into Christian thought. That is why disputants still often rely on it today. (Other Greek thinkers, such as Porphyry, who defended animals, were largely forgotten). For the Stoics, reason played a central part in morals. Reason was the sole basis of duty because it united all rational beings in a supranational community. As far as human life went, this was a noble concept which did much to discredit slavery. But it excluded irrational beings and allowed of nothing intermediate. Stoic theorists therefore argued determinedly that animals could not reason at all. This led them constantly to give strange and distorted definitions both of reason itself and of the various other faculties, such as perception and emotion, that might be brought in to do its work. They thus drove a gratuitous wedge between reason and the rest of the human personality, producing a wide gap that still gives trouble today.

Questions about animals are not, then, just an amusing extra. Their distortion has seriously harmed our thinking. Oversimplification of the moral issue is of course only one aspect of this distortion, but, as Sorabji says, it is one that can scarcely fail to strike us once we notice the confusions that have concealed it. We cannot today treat all our duties, Stoic-fashion, as pure tributes to the rationality of others, still less to their syntactical abilities. The first question we would now naturally ask is, as Jeremy Bentham rightly said, not 'can they speak?' or 'can they reason?' but 'can they suffer?' More widely, the Darwinian continuity that we now recognize between *Homo sapiens* and other species radically shifts the burden of proof from those who see possible connections to those who deny them *a priori*.

Sorabji does not offer any simply moral conclusions. He is rightly critical of the rather crude, sweeping 'moral theories' proposed by pro-animal philosophers such as Tom Regan and Peter Singer. But he is sure that the traditional policy of ignoring this issue is not compatible with the kind of moral principles we now live by. We do have to think the matter out afresh. And many scientists, I think, would now agree.

This is an impressive, important and thought-provoking book. It is clearly written and, although much of it deals with the details of Greek thought, the structure is plain and the index is excellent. □

Mary Midgley is at 1A Collingwood Terrace, Newcastle-upon-Tyne, NE2 2JP, UK.

On the edge

Malcolm W. Browne

A Positron Named Priscilla: Scientific Discovery at the Frontier. By Marcia Bartusiak et al. National Academy Press: 1994. Pp. 348. \$29.95.

ALL too often, omnibus science samplers intended for lay readers consist of breezy, easily digestible chapters on black holes, dinosaur lifestyles, insect navigation, dolphin communication, space travel, ecological issues and the like; they contain precious little to nourish the minds of people interested in the details of physical science, including (perish the thought) a few illustrative equations. *A Positron Named Priscilla* is a delightful exception, and (for the most part) a pleasure to read.

The whimsical title refers to an experiment conducted in the late 1980s by Hans Dehmelt at the University of Washington, who succeeded in trapping and immobilizing a single positron using electric and magnetic fields. As he conducted precise measurements of his caged blob of antimatter, Dehmelt named it Priscilla, explaining in his *Science* paper (as quoted in the book) that "the well-defined identity of this elementary particle is something fundamentally new, which deserves to be recognized by being given a name, just as pets are given names of persons".

Dehmelt's work is described in one of two excellent essays in this collection by T. A. Heppenheimer — the first on new laboratory implements (such as the scanning tunnelling microscope) that can manipulate single ions, neutral atoms and particles, and the second on the quest by particle physicists for the elusive top quark and the Higgs boson.

Heppenheimer, like the authors of the other eight essays in this collection, starts his discourse with some easily grasped ideas, from which he gently leads the reader into the thickets. The tour takes one through a fascinating labyrinth of interconnected research, historical turning points, personalities and problems, and makes it easier to place new scientific developments into perspective as they come along.

The quality of writing in these sophisticated essays varies; a few of the authors might have remembered that effective prose requires more than merely getting the facts right. A small problem with several chapters is the intrusion of jargon; 'difficult' becomes 'non-trivial' and a number is 'plugged in' rather than 'substituted'. Also, a few writers occasionally try to explain something by merely labelling it. Some of the biochemistry may glaze the eyes of readers confronting a dense alphabet soup of substances known by their three-letter labels — names familiar

to specialists but confusing to outsiders.

But most of the book is remarkably lucid. An example of science writing at its best is an essay by Barbara Burke on some of the mathematical tools used to extract signals from mountains of noise. She shows how these tools help to track submarines, catalogue fingerprints, sort out faint radio signals from celestial objects and even reconstruct the primitive and crumbling recording of an original piano performance by Brahms. From such examples she leads the reader into commendably clear explanations of Fourier analysis, and then to a new mathematical entity called the 'wavelet', which, in the hands of an adept, can supposedly "cut the weeds and spare the daisies".

Although many fields of research are ignored, the chosen subjects are compelling because all of them deal with work in progress. Each essay describes conspicuous accomplishments but leaves the disquieting impression that ultimate success is by no means assured.

An essay by Michelle Hoffman about the diabolical tricks devised by the chameleon-like HIV virus to overwhelm a person's immune system, points to some promising potential defences, but also explains why some scientists believe that a cure for AIDS may never be developed.

Indeed, some of the fields may already have hit major obstacles. The demise of the Superconducting Super Collider has dealt a heavy blow to hopes of identifying the Higgs mechanism — a theoretical pillar of the Standard Model and the putative universal giver of mass. And in earthquake forecasting, geologists and seismologists may have to accept that reliable, short-term predictions of earthquakes are intrinsically impossible. It may be that the timing of an earthquake is critically dependent on undetectably small differences in initial conditions — that in essence earthquakes are chaotic.

Other fields of research, enriched by new devices, techniques and discoveries, are booming. Marcia Bartusiak offers an exciting glimpse of the astonishing precision of Doppler interferometry in detecting the vibrational modes of the Sun, vibrations that reveal the deep, hidden mechanisms that power stars.

In chemistry, the discovery that carbon atoms can join together as the football-shaped molecules called fullerenes has created a field of investigation as potentially rich as the one that evolved from the discovery of the structure of benzene.

This is a breathtaking overview of some of the most exciting facets of science at the turn of the century, and there's gold to be mined here, whether the reader is a scientific specialist curious about alien fields of research or merely a science buff. □

Malcolm W. Browne is a science writer on The New York Times.

The making of American dreams

McGeorge Bundy

James B. Conant: Harvard to Hiroshima and the Making of the Nuclear Age. By James G. Hershberg. Knopf: 1993. Pp. 948. \$35.

PROBABLY no American in his generation made a larger mark in more different fields of intellectual activity than James B. Conant (1893 – 1978). He was a first-rate chemist, a great university president, a brave and honourable pioneer in nuclear weapons policy, a bold reformer of public education and second only to John J. McCloy as an American contributor to the new Germany. James G. Hershberg has now made a massive effort to provide the biography that has long been needed. His *James B. Conant* has important omissions and a few serious misjudgements, but it is much more often right than wrong. It is also a book so honest and comprehensive in its scholarship that the careful reader can rely on Hershberg's evidence even while doubting one or another of his conclusions.

What is most seriously missing from the book is a full explanation of what eventually defined Conant as a public man — his standing as the deeply respected and eminently successful president of Harvard University. Hershberg tells important parts of that story, such as the shift from Lawrence Lowell's focus on college life to Conant's emphasis on the professors and the crucial retreat to victory in an early battle with the central faculty over appointments. But he misses the main point: that Conant understood, exercised, defended and reinforced the final authority of the Harvard president to decide who is recommended for any faculty appointment to the governing boards (there are two at Harvard), and also who is recommended to be a provost or a dean, or indeed to hold any other office of administration. Conant exercised those powers with an integrity and intelligence that so reinforced the Harvard presidency that its basic powers have been maintained ever since. Its powerful presidency is the single most important advantage that Harvard has over its principal competitors in the front rank of American universities.

It was his standing as president of Harvard, together with his running start in scientific understanding, that made Conant the ideal choice to be the senior scientific watchman over the vast secret nuclear enterprise of the Second World War. It was also right for him to do that job while continuing as president in Cambridge, mainly as a weekend commuter on the night train between Boston and

Washington. "President Conant" was a person whose name and title were resonant both in recruitment of scientists from the academy and in counsel to the Harvard man in the White House. Hershberg notes but hardly pauses to explore the extraordinary partnership between Conant and Harvard provost Paul Buck, a relationship that made this wartime arrangement possible and continued for eight years beyond, making time for the later work of President Conant the Public Citizen.

The part of Conant's life that most interests Hershberg is his work on the problem of the nuclear bomb. About half



On the waterfront — James B. Conant in 1953.

of the book relates to this subject, and those who are more interested in Conant than in the bomb will find a lot to skip. The passages on great moments are riveting: Conant's decisive perception in 1941 that the bomb was possible; his experience of the first nuclear explosion at Alamogordo in 1945; his attempt to find a way to hold back from thermonuclear weapons in 1949; his courage and ineffectiveness in an effort to prevent the outrageous injustice inflicted on Robert Oppenheimer by finding him a security risk in the Eisenhower years. The attentive audience for other parts of his nuclear story may be smaller, but for specialists Hershberg has done a great service by this detailed account of the changing thoughts of a senior participant in and observer of the bomb's first quarter of a century.

In his discussion of Hiroshima, Hershberg also gives us a wonderfully clear illustration of a recurrent stress between

his professional respect for evidence and his left-of-centre political sympathies. The central question here is what principal motive governed the decision to bomb Hiroshima — to end the Japanese war or to get ahead of the dangerous Soviets? Hershberg has found out all there is to know, and with one part of his mind he understands the historical record of the decision very well. With another part he disapproves of it, and so he concludes that Conant's campaign to defend that decision was somehow unworthy. In a footnote on page 819, he recognizes that Conant's basic view of the reasons for the Hiroshima decision is now generally accepted by professional historians. He reports their "loose consensus" that the "principal (but not *exclusive*) goal" of the attack was "winning the war as quickly as possible with minimum loss of Allied lives". Still, he somehow finds it conspiratorial for Conant in 1946 to urge Henry Stimson, as the former Secretary of War

who had been the senior responsible official under Harry Truman, to defend the decision and then to help Stimson in preparing the resulting magazine article. (I know the story well because I was Stimson's assistant and scribe at the time.) Conant may have been basically right about the history, and even right that it should be explained, but Hershberg's underlying political sympathy is with later critics who have claimed that the main purpose was to impress the Russians, and he gives the chief exponent of that erroneous view the last word in his own account. The young scholar respects the claims of honest history and tries to honour them, but the young progressive gives an edge to the revisionist opinions that he would prefer to share.

I think there is a similar rush to judgement in Hershberg's assessment of Conant's role in Harvard University's resistance to McCarthyism. He believes, as I now do myself, that Conant went too far in his flat announcement that he would never knowingly join in the appointment of a Communist. Practically that judgement may have been sound, because the remaining members of the American Communist Party, from 1948 to 1953, were a generally sorry lot in terms of intelligence or intellectual honesty or both. Nevertheless, the rule was too broad. Hershberg also correctly notes a certain excess of caution in the Harvard Corporation's approach to ex-Communists. But his largest complaint is against Conant's opposition to the use of the Fifth Amendment — a witness's refusal to answer questions about other possible Communists on the ground that such answers could incriminate him or her. The historian in

Hershberg notes that Harvard in the end took a less stern position, which ex-President Conant supported. Whether it was right or wrong, Conant's initial view of resort to the Fifth Amendment was not a large matter in Harvard's overall struggle against McCarthyism.

It is a very long jump from this set of complaints to Hershberg's sweeping conclusion that on this front Conant "moved toward a subtle yet pernicious form of thought control". That is not the picture of Conant or his influence that those who were there remember, and neither Hershberg's own comments nor his footnote references give evidence supporting any such conclusion. The historian, once again, does not confirm the critic.

Diplomacy was not Conant's long suit. He failed conspicuously in wartime negotiations over nuclear cooperation with Britain, creating an impasse that was resolved only by Roosevelt, Churchill and Stimson. Nonetheless he very much wanted the appointment he won in 1953, leaving the Harvard presidency to succeed McCloy as the principal US representative in Germany, as High Commissioner and then as the first Ambassador. In these jobs, as Hershberg demonstrates, he was not so much unsuccessful as unnecessary. Dulles in Washington wanted no independent adviser in Bonn, and Adenauer recognized that he could do better by going directly to Dulles than by going through a man who did not know how to have no opinion of his own. Conant was neither the first nor the last ambassador to misunderstand the range of his role in the age of the airplane and telephone, but his appointment, and a later two years as a private citizen in Berlin, did allow him to play a significant personal part as an outstanding American who honoured the best of the German past, especially in science, and deeply believed in an honourable German future.

Finally, in his last round of public service, Conant became the leading advocate of reform in American public schooling. Here he made a mark exceptional in its force and clarity, but he would not be surprised that the American nonsystem has been resistant to his basic programme of reform: quality education for all children in public schools, to provide both equal opportunity and a fully demanding challenge to the young of all capabilities. Hershberg finds him farsighted in observing and describing the failure of public schooling in urban ghettos, but at first insensitive to the way black leaders would respond to his apparent acceptance of *de facto* segregation. Once again the analyst outran the diplomat. Conant's volumes on public education were nonetheless a major contribution, in yet another field, to understanding of a continuing American problem.

Conant was not an easy man to get close

to, and it is not surprising that Hershberg has trouble understanding his family life. He tries hard, and Conant's family have tried to help him, both with family papers and with interviews. But the closer he gets to the man, the less certain his touch. The primary cause of the trouble is Conant himself; he was a genuinely reticent descendant of the Pilgrims. Yet the best guide here is not Hershberg's book but Conant's memoir, *My Several Lives* (underrated by Hershberg). Conant had an essentially wonderful marriage, for example, and that reality shows through in his own book. Hershberg offers more extensive evidence on family troubles, but except for the grievous illness of the older son it is mostly about difficulties of the kind that come sooner or later to most families. There is no matching emphasis on the integrity and intensity of a great marriage, old-fashioned in that the man's work came first, but old-fashioned also in his reliance on his wife, and in the way they made most large choices together.

Overall, this book is an impressive achievement. In the modern process of American academic appointment and promotion that Conant did so much to

shape, nothing is more important than a young scholar's first major book. After Conant's time, I sat as a Harvard dean through dozens of presidential meetings in which such scholars and their books were considered, and I have found myself wondering how this book would fare before such a committee with Conant himself presiding. I think there would be criticism and perhaps even complaint from the chair, but Conant was an extraordinarily fair-minded man. He might well propose to disqualify himself, but I think such a proposal would not be in Hershberg's interest. I think that Conant would turn out to share my own strong opinion that on the evidence of this book the author has shown himself to be a historian of notable achievement and promise, and well qualified, as far as a single book can show it, to do with distinction the work that Conant himself may in the end have cared about most: the work of a permanent member of the faculty in a leading research university. □

McGeorge Bundy is in the Carnegie Corporation of New York, 437 Madison Avenue, New York, New York 10022, USA.

Hearing the deeper harmonies

Laura Garwin

The Hidden Landscape: A Journey into the Geological Past. By Richard Fortey. Cape: 1993. Pp. 288. £19.99.

I HAVE never met Richard Fortey, but from his book I can think of no better companion for a walk through the British countryside. Stand with him on a bleak Yorkshire moor and be transported to a lush rain forest of 100 million years before; gaze across the low, rounded hills of northwest Scotland and recognize a landscape inherited from the Precambrian; look down on the waters of the English Lake District and see reflections of a glacier that vanished 7,000 years ago. But these links between past times and present landscapes are just the start, for Fortey's mission is to show us that the influence of the rocks beneath our feet extends far beyond scenery — to flora and fauna, agriculture and architecture, and beyond, into our daily lives.

The 'hidden landscape' of the book's title has several different meanings, the most straightforward of which refers to the fact that the bedrock geology of the British Isles (and indeed of most parts of the world) lies hidden beneath a blanket of soil and vegetation. Fortey imagines himself "rolling back this skin, as one might a carpet, to reveal the rocks beneath". Applying this process to the British Isles reveals an extraordinary variety

packed into a small space: sedimentary, igneous and metamorphic rocks, and representatives of each of the geological periods. This richness of geology is one way in which the book avoids being too parochial. The other is Fortey's approach: by travelling chronologically upwards through the geological column, rather than geographically around the British Isles, he can tell us something of the history of life on Earth, and of the wandering continents, along the way.

In fact, Fortey packs a lot of modern Earth science into the book — including the story of the vanished ocean of Iapetus, which separated northern from southern Britain until it closed about 400 million years ago, leaving only the Solway Firth, the sediments of the Southern Uplands, and some otherwise very confusing fossils to bear witness to its former existence. We also learn about radiometric dating, experimental petrology and even the early Solar System. But while these forays into geological science undoubtedly enrich the broth, Fortey's greatest delight lies in revealing the connections between geology and our everyday experience of the natural world.

And well might he delight, for the connections are many and multifarious. At the broadest, most familiar scale, we see how the Highland Boundary Fault creates a sharp transition between the comfortable farmlands of the Midland

Valley of Scotland, which overlies sedimentary rocks, and the wild highlands to the north, betraying the presence of tougher metamorphic rocks beneath. But Fortey tells us, too, about the plants that flourish on the nutrient-poor moors of the highlands: for example, the berries of the heath family (Ericaceae), which spark a paragraph's digression about the problem of obtaining vitamin C in high-latitude regions. Other highland plants prompt reflections on the survival of primitive forms in harsh conditions (the clubmosses on today's moors trace their heritage back to Carboniferous coal swamps) and the stranding of Arctic flora on British peaks as the glaciers retreated.

From flora, on to fauna: as the heather follows the geology, so does the grouse. Why are brown trout found in the streams of both Derbyshire and Hampshire? Because both regions are underlain by limestone — respectively, Carboniferous and Cretaceous (the famous Chalk). Nor are humans impervious to the influence of the hidden landscape. Fortey goes beyond the obvious examples of agriculture (dictated by the fertility of the soil, itself controlled by the bedrock) and mining, to argue that much of today's human geography still reflects the geological imperative. The vales flooded by Permian and Triassic rocks are "natural corridors for traffic and trade"; hence one finds canals along them, and also ancient towns. The wool towns of western and middle England grew up near the outcrop of the Fuller's Earth — a Jurassic deposit rich in the clay mineral montmorillonite, which can be used to clean wool of its natural grease. The modern chemical company ICI owes its location, argues Fortey, to the outcrop of Permo-Triassic salt.

To appreciate many of Fortey's examples, one needs to be fairly familiar with British geography. Instead of having the geological map of Britain and Ireland reproduced twice, could we not have had a topographic map, to aid readers who would not otherwise know what it means for a rock formation to extend "from Nottingham to Sunderland, with its widest outcrop at the coast north of Hartlepool"? More generally, one often gets the feeling that Fortey is writing for a British audience — which is a shame, as this is emphatically not just a book about the geology of the British Isles; it is about how an appreciation of geology and its hidden connections can enrich one's experience of life in exactly the same way as can an understanding of art or music. To quote one of Fortey's more whimsical examples, does it not deepen the enjoyment of a whisky and mineral water to notice that although one ingredient is Scottish and the other English, they both originate in rocks of the Precambrian era?

But the ability to recognize harmony necessarily brings with it a sensitivity to

discord. The flip side of the pleasure provided by vernacular architecture — buildings made from local stone, "at ease with [their] geological foundations" — is the pain brought on when the connections are ignored. In this regard, Fortey is perhaps an extreme geological aesthete: he rails, not just against the brightly coloured petrol station which we might all recognize as a modern synthetic abomination (although he adds that it does not even have the modesty to clothe itself in lichens, as the local stone does), but against the embarrassment of houses made of Jurassic Bath stone sitting on the Cretaceous chalklands. One feels for him, as for the person whose hyper-perfect pitch ruins the enjoyment of an average concert; but even those of us with more normal sensibilities will acknowledge the loss entailed in the march of uniform brick and slate across what was a richly varied landscape.

Fortey's stated aim is to "inspire a way of looking at the landscape". Read this book and you will never see the world around you in quite the same way again. □

Laura Garwin is Physical Sciences Editor of Nature.

Technopolis

Henry Petroski

A Scientist in the City. By James Trefil. *Doubleday: 1994. Pp. 266. \$23.95.*

AMONG the fundamental differences between scientists and engineers are their generally divergent objectives. The purest scientists observe and study the given world only to understand and explain it and its phenomena; the purest engineers want to manipulate the parts and processes of the given and found world into artefacts of benefit to mankind. Scientists analyse; engineers synthesize. The products of science thus tend to be words and theories; the products of engineers are plans and things.

In his latest book, the scientist-writer James Trefil turns his analytical eye to the quintessential products of synthesis — cities and their artefacts. As with much science, his musings begin in wonder, specifically atop the Sears Tower looking down at Chicago, and he progresses from wondering about what problems had to be solved to build the world's tallest building to wondering "what else in the city depended in a similar way on our basic understanding of science and nature".

Allowing Trefil the simplistic conventional wisdom that engineering and technological advances are little more than applied science, the premise of his book is an excellent one, and even an engineer

might sympathize with the popular goal of wanting the reader to see modern cities as "products of a series of discoveries of the physical universe". Trefil proposes in his introduction to accomplish this objective by looking at three discoveries: "the ability to manipulate atoms"; "the ability to unlock stored energy"; and "the ability to store and transmit information electrically". But the reader might be advised to keep in mind that each of these discoveries or accomplishments is in fact a feat of engineering that may or may not involve science in the usual sense. Indeed, manipulating atoms over fire, perhaps one of the essential hallmarks of civilization, required neither formal science nor engineering, although the craft may have been honed by the rudiments of their methods.

Despite oversimplifications of fundamentals, which are understandably made in the interest of getting on with the story, the book is a good read. Trefil tells interesting tales about the development of such materials as the glass and steel that go into modern skyscrapers, and he explains well the problems associated with moving things such as people, water and waste in very tall buildings. The first half deals with the city of today, and with Trefil's personal encounters among the systems of several specific cities; the second half of the book deals with the city of the future. Amid discussions of the implications of life among very tall buildings and the nature of suburbs, Trefil covers a lot of ground, sometimes as fast as the Maglev trains that he describes as redefining the distances commuters of the future will tolerate, in the rare time they might have to leave the information highway at home and travel to the office.

In the end, the success of science and engineering, and of books on them, rests in the details. Trefil assures the reader that "engineers, you know, are a gloomy lot, always dwelling on what can go wrong with their systems", but he neglects to remind us that by "dwelling on what can go wrong", engineers are striving to get it right so their products can be relied on by those who are not engineers. Who wants engineers to be so sanguine about an elevated walkway that they leave the design details to the construction workers?

Although Trefil is not an engineer, he might have been a bit more critical of some of the details on which his own construction rests. For example, what he says engineers call "shear loads" they might hardly recognize. Steel reinforcing rods do not provide strength directly against shear, as Trefil states on page 36, but against tension, in which concrete is weak. The illustrations in the book can be equally wrong. The figure purporting to show the "flow of force on a lintel", for example, has the lintel between the posts rather than resting on top of them. The

caption to a figure comparing Frank Lloyd Wright's mile-high skyscraper (which was never built) to more familiar structures states that they are drawn to scale, and yet they are not. Such a neglect of detail jeopardizes an otherwise intriguing book, for the gloomy engineer has naturally to wonder about the details of those parts for which he must rely on the scientist. □

Henry Petroski is in the Department of Civil and Environmental Engineering, Duke University, Box 90287, Durham, North Carolina 27708-0287, USA.

Behind the birdies

Stephen Salter

The Physics of Golf. By Theodore P. Jorgensen. American Institute of Physics/Oxford University Press: 1993. Pp. 155. \$35, £27.50.

It was on grounds of national security that the Scottish Parliament introduced successive acts (1451, 1457 and 1491) banning the game of golf. The unprofitable use of land, the loss of productive working hours, the stresses on family life, the danger of apoplexy and the reduced attendance at kirk were factors but it was the distraction from archery practice that was decisive.

With advances in military technology, golf has been decriminalized in Scotland and many other countries. Golfing equipment with a street value estimated at \$13 billion a year is now sold quite openly. Jorgensen's book on the physics of golf offers no suggestions for treatment of addicts but may help us to understand their problems.

His model of the golf swing consists of two stiff hinged rods, one representing the arms of the player and the other the club. He presents the algebra to predict the behaviour of this model in terms of five torques on the golfer's arms. He supports the predictions with stroboscopic photographs of patches of reflective tape stuck to the club and to points on the body of the golfer. The observations are in good agreement despite the simplicity of the rigid two-rod model.

Many of the results are intriguing. For example, a marker on the head of a golfer will trace a curve that is as unique as his signature — a more enjoyable method of security identification than fingerprinting. In a good golf stroke the velocity of the golfer's hands decreases on contact with the ball, while that of the club head continues to increase. Similarly, a bull

whip makes a supersonic crack when most of it is static and all the kinetic energy has been transferred to the very tip.

Did you know that bull whips should have lead weights near the handle? Did you know that you can break the shaft of a club by not allowing your wrists to rotate? Did you know that a perfect ball would leave the tee at twice the speed of the club head? Good players achieve a ratio of 1.4 which suggests that there may still be room for design improvement.

Jorgensen calculates the instantaneous power delivery of human muscle and says that the amount of muscle needed (32 pounds or 15 kilograms) is far more than is available in arms and shoulders (he suggests that a visit to the butcher will confirm this). He argues that other muscles in the legs and back must be in use, passing energy through the arms.

The book contains enough experimental detail to allow readers to replicate results. An interesting example is the work of Peter Guthrie Tait in 1887 on ball spin. Working without electronic instruments or fast photography, he measured spin by attaching a flat ribbon to a golf ball and driving it into a nearby bank of thick clay. By counting the twists of ribbon, Tait calculated that the rate of spin was as much as 120 turns per second or 7,200 revolutions per minute, depending on the angle of the club face. Many people did not believe him.



Swinging both ways — woman golfer from *La Vie Parisienne*, 1908.

The rotation is produced in the fraction of a millisecond in which the ball is in contact with the club face. It has a profound influence on the flight of the ball because of the Magnus effect. The author explains how it can induce a lift force nearly as large as the gravitational weight of the ball.

Jorgensen explains how to match a set of clubs so that they all feel the same

despite the different weights of their heads and the fact that golfers like clubs of different lengths in their set. The method involves placing small weights at two places along the club shaft so that the total weight, the first moment and the second moment are the same. Small changes of weight, of the order of 1 gram, are involved. Good golfers seem to like the results.

A whole chapter is devoted to the laws of probability and their relationship to golf. The author collects data from the score cards of a pair of players over 20 games and calculates the probability of various scores. The Gaussian distribution strikes yet again. Twenty games — about 1,500 strokes — seem to be plenty to give the familiar smooth curves. Golf authorities have not so far devised a way of awarding non-integral handicaps. Jorgensen shows that this will prevent better players retaining an advantage.

Jorgensen writes for a wide readership ranging from nongolfing fellow physicists, who may find some of the verbal explanations rather laboured, to nonscientific golfers, who may wince at the algebra. He spans the gap with appendices for the heavier maths; readers should be tolerant.

Relaxed sportsmen in the United States will be less irritated than pedantic academics in Europe by the promiscuous mixture of physical units. Many results are given in inches, feet and yards with ounces and pounds for force. When Jorgensen goes metric he uses grams and centimetres. Club makers use an extraordinary measure, the Lorythmic swing-weight scale, in which ounce-inches of moment about a point 12 inches from the grip end are divided into letter numbers in bands of 20 ounce-inches, which are then subdivided into numbered bands of 2 ounce-inches. We can be grateful to Jorgensen for a clear explanation.

It was rather mean of the publishers not to include one of Harold Edgerton's splendid flash photographs of a golf ball flattened to half its diameter: the peak force between club and ball can exceed 3,000 pounds — more than the weight of a family car. There is only one page of references and a three-page index, but this will not detract from one's enjoyment of the book.

Golfing addicts now need a psychologist to explain the ecstasy of hitting a golf ball exactly right. I can testify to the addictive power, managing such a hit once in 1958 and again in 1987. No wonder the Scottish Parliament needed more than one act. □

Stephen Salter is in the Department of Mechanical Engineering, University of Edinburgh, Edinburgh EH9 3JL, UK.

Through the doors of deception?

Charles Medawar

Listening to Prozac. By Peter D. Kramer. Viking: 1993. Pp. 409. \$22.50. To be published in the United Kingdom by Fourth Estate on 7 April at £16.99.

If all goes well I will write an essay on it and I expect it will win its place in therapeutics, by the side of morphium and superior to it. I have other hopes and intentions about it. I take very small doses of it regularly against depression . . . and with the most brilliant success.

FREUD always denied that his enthusiastic advocacy for "this magical drug" was ill-judged, but his peers soon concluded otherwise and he was roundly criticized for it. Freud not only believed that cocaine was a miracle drug; he had also hoped that by revealing this truth he would secure professional advancement and wealth. At first he saw no risk of addiction to cocaine; later he reasoned that any such problem would signal some personality problem rather than any defect of the drug. Accordingly, the young Freud pressed cocaine on patients, family, colleagues and friends, persuading even his official biographer that, "looked at from the vantage point of our present knowledge, he was rapidly becoming a public menace".

Years before this Freudian skid — and ever since — doctors have prescribed a virtually uninterrupted succession of addictive drugs for anxiety and the like, each time in the mistaken belief that the drugs would not cause dependence or that it was the patient's fault if they did. For years, morphine was used to treat opium addiction, and heroin was later routinely used to treat addiction to cocaine. This list is long: it now includes numerous barbiturates and related drugs and extends beyond even the benzodiazepines (such as Valium and Librium), still widely prescribed today.

It took more than 25 years to establish how addictive the benzodiazepines really are. Even a decade after the introduction of Librium, its manufacturers could get away with a centre-spread advertisement in a learned journal headlined: "Whatever the diagnosis — Librium" (*British Medical Journal*, 1 March 1969). And there were positive streams of anecdotal clinical evidence to prove it:

The effect of Librium on the symptom of excessive withdrawal and tension (excessive control) is remarkable, and may be illustrated by the description volunteered by one patient, a teacher: "I think it is miraculous. I was feeling outgoing. The effect was so liberating that I felt excited and very outgoing . . . I would walk into the class completely unprepared and end up well. It made me become less involved with myself and my state of mind was outgoing" [Toll, *N. Disorders of the Nervous System*, 264–6, March 1960].

My reading of Dr Peter Kramer's book about Prozac — the antidepressant drug fluoxetine — has been much influenced by this historical background. So far as one can tell from controlled clinical trials, Prozac is about as effective as other antidepressant drugs, yet Kramer claims that with many of his patients he has had sensational results. In the United States,



Hyped-up — Prozac makes the headlines.

the book has become a huge bestseller and Prozac is now a craze; some five million Americans take the drug and the numbers are rising. Working quite independently of the manufacturers, Kramer must have sold far more Prozac with his book than they could have ever hoped to achieve alone: sales now exceed \$1.2 billion per year.

Kramer's book is based on anecdotes, most of which detail the experience of about a dozen of his patients; they found not so much that Prozac cured them of illness but that it transformed their lives. The word 'health' is not to be found in the index; this book is about the alleged propensity of fluoxetine to produce in some unspecified proportion of fairly

healthy users increased faith in themselves; greater hope for professional, social and personal growth; and love of life. In a nutshell: "since you only live once, why not do it as a blonde? Why not as a peppy blonde?"

The explanation for the intriguing title is Kramer's view that fluoxetine can be used both to understand and then to shape traits of character, perhaps achieving at a stroke what followers of Freud might take years to achieve. But make no mistake, this book is not about "Listening to Prozac" but about catching the drift of a somewhat beamed-up psychiatrist. Kramer's statement of his credentials hints of keen identification with the world of *Star Trek* and Captain Kirk:

My quest has led me deep into territories whose results inform my clinical work but whose culture and customs are foreign: cellular physiology, pharmacology, history of medicine, animal ethology, medical ethics, descriptive psychiatry" [p.xix].

And this seems a heavy boast considering the lightweight quality and style of what Kramer really has to offer. The nub of his evidence on the effects of Prozac on his patients' lives is not science but soap:

"Three dates a weekend," Tess told me. "I must be wearing a sign on my forehead!" Within weeks of starting Prozac, Tess settled into a satisfying dating routine with men" [p.7].

Now she strode forward and gave me a bold "Hello." I responded, and she said, "I've changed my name, you know." . . . She had, I saw, the bright and open manner that had brought Tess so much social success. "Yes," she continued, "I call myself Ms. Prozac" [pp. 11–12].

"People on the sidewalk ask me for directions!": I have since heard this identical report from other people on Prozac [p.320].

This is the content and tone that have made the book a commercial success, although its credibility has been bolstered with a fair amount of more baffling material besides. Psychobabble accounts for part of it and Kramer also makes a determined attempt to explain some of the very little that is known about the chemical niceties of brain function in the regulation of mood and mind.

Moral and social issues are raised too, and centrally this question: is it ethical to treat someone who is not ill with a drug that makes them feel "better than well"? This, to my mind, is alarmingly disingenuous: Kramer clearly has miraculous transformations in mind, and proudly coins the term "cosmetic psychopharmacology" to prove it. He must have known that to write so seductively about a feel-good drug would stimulate mass demand and inevitably some misuse.

We do not yet know enough about this drug, and it will be years before we do. Fluoxetine was tested in controlled trials

lasting only a few weeks, and was licensed only a few years ago on the strength of results that gave no hint of the easy trips to which Kramer refers. (I would not be surprised if some passages evoked in cannabis users a sense of *déjà vu*.)

It certainly will not do to declare: "It is not addictive — patients do not crave Prozac, and there is no known withdrawal syndrome" (p. 311). The clear message of history is to beware of any explosive, mass demand for a psychoactive drug, and never to forget that patients don't crave so long as doctors readily prescribe. It is also worth noting that there is still profound confusion over the differences in meanings of 'dependence' and 'addiction', despite clarification from the World Health Organization 30 years ago. The Royal Colleges of Psychiatrists and General Practitioners still insist that "antidepressants are not addictive" and that people are "mistaken" in thinking they can cause dependence (RCP/RCGP Defeat Depression Campaign Release, 1992), despite evidence that most antidepressants (unlike cocaine) are associated with a withdrawal syndrome. This is why the *British National Formulary* recommends that when patients stop taking these drugs, "reduction in dosage should preferably be carried out gradually over a period of about four weeks".

What the Royal Colleges are really saying is that antidepressants have no great market street value and do not lead to overt drug-seeking behaviour. But this does not properly address the concern that both doctors and patients may misinterpret withdrawal symptoms as recrudescence of disease, and then see this as evidence of the effectiveness of the drug and of the need to continue treatment with it. This is what happened with the benzodiazepines, barbiturates and all the rest, and it led to dependence on a grand scale.

And the risk would be greater with drugs such as fluoxetine, which are only slowly cleared from the body. This means that withdrawal symptoms tend to peak long after the drug is stopped and therefore that the drug would be less likely to be identified as their cause. In any case, if one defines 'dependence' simply as 'drug-induced drug consumption', one would never conclude from Kramer's evidence that fluoxetine is a drug that anyone might take or leave at will:

We lowered the dose of medicine, and two weeks later Julia called to say that bottom had fallen out: "I'm a witch again." She felt lousy — pessimistic, angry, demanding . . . and then she used the very words Tess had used: "I don't feel myself." . . . Julia resumed taking the higher dose of Prozac. Within two weeks, she felt somewhat better; after five weeks, she was "almost there again," with many more good days than bad. She said work had been torture on the lower dose of medicine [pp. 29–30].

Bearing in mind that "Prozac has not been systematically studied, in animals or humans, for its potential for abuse, tolerance, or physical dependence" (*Physicians Desk Reference*, 4th edn, 943–6 (Medical Economics, Oradell, New Jersey, 1993)), it would be folly not to undertake searching enquiries into how Prozac is used and misused and to what effect. Some users clearly do benefit, although in pre-marketing trials one in seven patients had to quit because of adverse effects, most commonly psychiatric (*ibid.*). The broad spectrum of these unwanted symptoms is worrying, with some people becoming badly agitated and hyped-up, and others brought right down; in extreme cases, manic psychoses and suicide have been mentioned as real risks. In a mass market there will certainly be many casualties and much more could be done to prevent them.

Then there are uncertainties about dosing. Noting that patients responded in the wide range of 5–80 mg per day, the Swedish and Norwegian authorities in 1991 refused to give Prozac a licence —

because it was supplied only in a 20-mg size. Kramer has little to say about doses, and where he does he seems unduly relaxed:

[Gail] asked whether I could raise the dose of Prozac so she would feel comfortable applying for the post. I did not know whether a different dose would have a different effect, but I saw no reason not to try. She took extra Prozac, and she applied for the promotion. She was eventually turned down, but she was able to take the rebuff in stride [p.94].

Who knows, Prozac may turn out to be the triumph of benefit over risk. In the meantime, it might be safer to regard this book as an object lesson in how not to evaluate a drug, and as a dire if inadvertent warning that products such as Prozac threaten to consume us all. We must not forget that the greatest mistakes in medicine tend to be made not because doctors don't know enough, but when they behave as if they do. □

Charles Medawar is at Social Audit Ltd, PO Box 111, London NW1 8XG, UK.

Ne'er the twain shall meet

Ryan J. Huxtable

The Billion Dollar Molecule: One Company's Quest for the Perfect Drug. By Barry Werth. *Simon and Schuster: 1994.* Pp. 445. \$25.

MUSIC lessons in my high school usually consisted of chanting "R-H-Y-T-H-M spells rhythm" or "Every Good Boy Deserves Fruit". On the rare occasions we had actual music, it was always something in a minor key, such as the song 'Behold a giant am I, aloft here in my tower'. This song is emblematic of the traditional academic, renowned in his field, living a monkish existence above the temporalities of common humanity. But, as another song has it, the times they are a-changing. Knowledge in US universities is now valued not for such unquantifiable attributes as 'for its own sake' but for the market price it carries. Forget the leather elbow patches on the smoky jacket. In the bright new world of managed scholarship and technology-transfer offices, the leather is on the briefcase of the professor raising venture capital for his biotech company.

This is the story of the collusions and clashes between industrial and academic scientists involved in a small company's search for a better immunosuppressant. A drug that can dislocate the biological system that asserts identity, distinguishing 'me' from 'not-me', is a valued product in an alchemical world of organ transplants. Hearts, lungs, bits of brains and even

complete viscera are, in the argot of the trade, harvested from one individual and incorporated into another. The surgery is straightforward. Transplantation, however, is vitiated by tissue rejection in the recipient. The alkahest that dissolves identity between donor and recipient, allowing the transplant to remain unrecognized, is an immunosuppressant. One immunosuppressant, cyclosporin, was isolated from a fungus growing in the wastes of arctic Norway. Another, FK-506, came from a mountainside in Japan. Both have numerous side effects. But suppose you engineered a drug specifically for the receptor that these immunosuppressants act on...?

The classic way of developing a drug once was to go to nature. That gave us morphine, vincristine, digitalis and taxol. Then the Faustian hubris of our species led us to think that whatever nature could do, we could do better. So we synthesized thousands of compounds more or less at random and tested them for drug activity. That gave us sulphonamides, INH, Salvarsan and propranolol. Then came Joshua Boger, founder of Vertex. His company would use a third, new way, of developing drugs: it would design them. Structure-based design would identify what a drug had to do and where it had to bind, and then the appropriate molecule could be custom designed. Line the balls up, tap gently with the cue and bingo! Illimitable largesse — the billion dollars of the book's title — would pour into the

pockets of the company. Midas never had it so good. There is a bright new language for this brave new world, such as "God mode" to mean total control, as in "we're in God mode".

Boger assembled a team of bright minds from Harvard and Merck and elsewhere. But the company was snookered before sinking the 8 ball. The coalition of magicians coaxed from academia or seduced from industry turned out to be illusory as the glass walls enclosing Marcel Marceau. The academics go their own way. The betrayed parent companies fight back with their own research programmes. Vertex hides information from its academic consultants. The consultants hide information from the company. Harvard gets concerned that it's losing control over profitable technology. The company continues to haemorrhage money like a politician up for re-election. It is rescued temporarily by white knights with Japanese faces, in the form of the Chugai company. Apollo is transformed into Dionysus as Vertex swerves into tantrums, fights, sleepless nights, drunkenness and traffic accidents. Success of a sort, however, does come in the fourth act of the drama. But ironically it comes by the method Vertex set out to eschew; the serial synthesis of compound after compound, ringing the changes on the carillon of structure. The three-hundred-and-sixty-seventh compound tested, V-367, has activity; here are the echoes of the Dengel's compound D-365 (verapamil) or Ehrlich's first success with the antispasmodic compound 606 (Salvarsan).

Illusion is the motif of this story. It

impresses by the amount of show rather than substance. From the molecules that exist on computer screens but not in test tubes, to the expectations of investors that correspond to no reality in the world of dimensions, to the negotiations with Chugai in Japan and Glaxo in England, this is a world in which honesty and straightforwardness seem as out of place as gonorrhoea in a nunnery. It is all flickering shadows in Plato's cave with no fire at the entrance. The only reality is the perception.

The book is up-to-date, covering events of barely a year ago. It is so up-to-date, in fact, that the fifth act, the resolution of the drama, is missing. The reader is left hanging like a car on an earthquake-shattered freeway. The storyline is not clear, and time, as befits the world behind the looking glass, does not adhere to chronology. The book, however, rattles along. During the unravelling of the story, the reader will learn a fair bit of the history of drug discovery and the rise of the pharmaceutical industry in this century. Many familiar names such as L. Pauling and R. B. Woodward make cameo appearances. Appropriate to our times, the story is an ambiguous one. Is it a story of success or failure? Or something in between? Is it a morality tale, and should "East is east, and west is west, And ne'er the twain shall meet" apply to academics and their journeyings to the industrial orient?

Ryan J. Huxtable is in the Department of Pharmacology, University of Arizona, Tucson, Arizona 85724, USA.

Hidden enemies, foreign invaders

W. F. Bynum

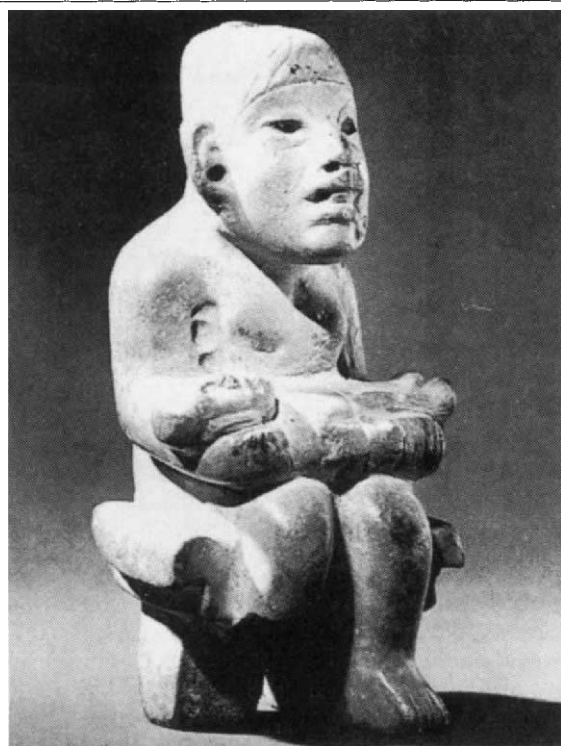
Silent Travelers: Germs, Genes, and the Immigrant Menace. By Alan M. Kraut. BasicBooks: 1994. Pp. 369. \$25.

Living in the Shadow of Death: Tuberculosis and the Social Experience of Illness in American History. By Sheila M. Rothman. BasicBooks: 1994. Pp. 319. \$25.

MY forebears crossed the Atlantic, probably from Holland, sometime before 1820, when the family genealogy falls silent. Most of Alan Kraut's immigrants went to the New World later, and from less prosperous countries: Ireland, southern Italy, eastern Europe, the Far East. The great wave of immigration spanned the period from the Irish potato famine of the 1840s, through the pogroms in eastern Europe of the late nineteenth century, the continuing poverty of Asia and the Mediterranean basin, and the disruptions of the Great War. Most emigrants were poor, taking little with them but their ambitions, language, traditions, genes and germs. For most, the language was not English.

The Great Melting Pot theme in American history has often been recounted, though not before with Kraut's medical slant. *Silent Travelers* weaves together two stories — the medical inspection that greeted would-be Americans on arrival, and the health problems that faced those who were allowed in. His subtext is the increasing concern of those born in the United States with foreigners, fears which led to tough immigration laws in the 1920s and made it difficult even during the Nazi period for refugees to count on the United States to become home ("the place where, when you have to go there, they have to take you in").

Periodical medical inspection of travellers and migrants has been around for a very long time, forming the administrative rationale for quarantine since the Middle Ages. It gradually became ritualized during the nineteenth century, and the depot on Ellis Island, New York, opened in 1894, was where millions of immigrants first encountered New World bureaucracy. Medical officers sought to identify those with dangerous or 'loathsome' diseases, the insane and 'idiots', and individuals thought likely to become a public charge. The rhetoric was pretty Draconian and Ellis Island figured prominently in the memories and memoirs of many who passed through. Nevertheless, the doctors made little impact on the immigration itself, never rejecting more than three per cent of those they examined. Despite the long list of conditions



SMALL stone figure of a woman and child from the Olmec, the most ancient Mexican civilization (Middle Preclassic, 1200–400 bc; 11.4 cm). The Olmec style is typified by small jade sculptures and other objects that emphasize human infants with snarling, jaguar-like features. It takes its name from the 'Olmeca', the mysterious 'rubber people' described by Sahagún and his Aztec informants as inhabiting the jungle country of the Gulf coast, to which many of these artefacts can be traced. In fact, nothing is known about the originators of Olmec art. This photo appears in the fourth edition of *Mexico* by Michael D. Coe (Thames and Hudson, £8.95 (pbk)), widely hailed as the classic introduction to the region's ancient civilizations.

that could exclude entry, only trachoma and favus were diagnosed with much regularity.

Kraut's second theme deals with the majority who got through Ellis Island and the other port screening centres. Except for a chapter on the Chinese in San Francisco, Kraut's focus is principally New York City, a quarter of whose population in 1860 had been born in Ireland. Half a century later, a quarter of the city's population was Jewish, and mostly first generation. Those who were not Jewish or Irish often came from Italy. Kraut analyses some of the health problems faced by each of these groups: exploitation in the workplace, poverty, overcrowding and insanitary living conditions, inadequate access to medical and

1910, but a few pages later are presented as the reverse in 1890. Why? The answer probably has to do with the reliability of the figures rather than the living conditions or health care of New York Italians, but we are never told. Throughout the volume, statistics are there to illustrate the social picture rather than to provide the starting point for analysis. Nevertheless, Kraut has a sharp eye for poignant detail and brings to life the individual stories of a few of the millions who made the journey.

In the 1903 *Book of Instructions for the Medical Inspection of Immigrants*, tuberculosis was a "dangerous contagious disease". Available methods of detection meant that few people were turned away because of it, but tuberculosis rates were high among immigrants and first-generation urban Americans. Some of these feature in Sheila Rothman's moving account of what it was like to live in the shadow of death. By the early twentieth century, Robert Koch's discovery of the tubercle bacillus had transformed perceptions of the disease from a constitutional, familial one to a contagious disorder best dealt with through isolation. In the transformation from consumption to tuberculosis, invalids became patients.

Rothman structures her narrative into three parts. The first, from 1800 to about 1870, deals with middle-class New England consumptives coming to terms with the probability of a shortened life of invalidism. Medical advice and conventional wisdom led the men to seek recovery through a sea voyage or wintering in warmer climates. Moderate exercise and fresh air were also deemed good for

weakened lungs. A few women also went along the same path, but travel for women with domestic responsibilities was difficult. Rothman captures the varieties of the male experience through the diaries and correspondence of many young men; the female is encapsulated through the brilliantly constructed story of Deborah Vinal Fiske (1806-44), whose confidante and fellow sufferer Harriet Webster Fowler followed her to the grave six weeks later.

Fiske's daughter, Helen Fiske Hunt Jackson, lived longer than her mother but she, too, was consumptive. She died in Colorado, where, along with Texas, Arizona, California and other western states, tubercular health-seekers settled in large

numbers after the American Civil War. They were generally tended by doctors who had gone west for the same reason: consumptives caring for consumptives. The promotional literature encouraged those with weak lungs to head west: "the Texas air I think beats any fattening compound I have tried for the last ten years", wrote Dr Boyd Cornick, who eventually set up his medical practice in San Angelo, Texas, a few miles from where my ancestors ultimately drifted.

After the voyage, and the west, came the sanatorium, as tuberculosis became a public health issue. The sanatorium movement incorporated many of the older notions of the relationship between climate and lung disease, but the new perception of the contagiousness of tuberculosis transformed social attitudes and underlay new medical powers of commitment. By 1925, there were more than 650,000 sanatoria beds in the United States, many funded by charities or municipal health budgets for immigrants.

Eminent doctors or reformers occasionally surface in these two books: Maurice Fishberg, Jacob Riis, Alfred Loomis, Edward Trudeau. For the most part, however, Kraut and Rothman seek to resurrect history's silent majority, to convey what it was like on Ellis Island or in the Adirondack Cottage Sanatorium. Inevitably, perhaps, each volume ends with the historical reverberations of AIDS and the salient reminder that meliorism is a hollow philosophy in 1994. □

W. F. Bynum is in the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

High-tech riches

Daniel S. Greenberg

Profits of Science: The American Marriage of Business and Technology. By Robert Teitelman. BasicBooks: 1994. Pp. 258. £23.

IN the 1960s, Wall Street called it the "sonics and tronics" boom — the rush for shares in companies radiating prospects of mammoth returns from the new gold fields of the United States, laboratories in the service of electronics and its associated industries.

Some early arrivals on the industrial scene, such as Texas Instruments, Xerox and Digital Equipment Corporation (DEC), fulfilled the high-tech investment dream, at least for many years, and others followed, such as Apple, Microsoft and Intel. Within a decade, DEC's initial backers reaped a 5,700 per cent return on investment. Unable to compete with this new way of creating wealth, some of the

NATURE • VOL 368 • 24 MARCH 1994

From Silent Travelers



Tuberculosis sucking the breath from a Jewish tailor's body (*Sanatorium*, 1907).

nursing services, anxiety and disillusionment (almost half the Italian immigrants ultimately returned to Europe). He examines strategies taken by and for new Americans to cope with disease, from imported traditional remedies to hospitals and health visitors.

Immigrant groups usually suffered excessive morbidity and mortality when compared with those born in the United States. Kraut's volume is littered with figures and the occasional table, but he seems less interested in trends than in snapshots of particular mortalities at random times. Sometimes the numbers cry out for probing, as when Italians are quoted as having much lower mortality rates than native-born New Yorkers in

famous old names of industry disappeared in great upheavals of money, marketing and technology, among them DuMont and Philco, powerhouses from the early days of radio. Mighty RCA, the giant of television innovation and sales, sold out to General Electric.

Two decades later there seemed to be a replay of corporate birth and growth in the life sciences, as investors responded to reports of an onrush in biotechnology that would even dwarf the spectacular triumph of electronics. "Biomania" was the term for this marketplace phenomenon, highlighted in 1980 by the initial public offering of shares in Genentech, a company with great promise but virtually no income. Genentech came on the market at \$35 a share and 20 minutes later had ascended to \$89. A decade later, Genentech, no longer the darling of Wall Street, entered into a corporate relationship with Hoffmann-La Roche to finance the difficult quest for profitability from biotechnology. Other companies in the biotechnology sector also soared in share price at the start, and subsequently went out of business or sold out to the established powers of the pharmaceutical industry.

The dynamics of these successes and failures is the subject of *Profits of Science*, an ambitious study that confronts the elusive relationship of finance and science and technology in the new postwar industries. Robert Teitelman, a senior editor at *Institutional Investor* magazine, is a knowledgeable guide through the finances and technologies of the centrepieces of this account: electronics, pharmaceuticals and biotechnology. Along the way, he also reaches for an understanding of the contributing roles of scientific progress, entrepreneurship, government policies and Wall Street's herd behaviour. Relying heavily on published sources, both scholarly and journalistic, this is a lucidly written work of interpretation, though in the end the author confesses his own puzzlement about many aspects of the dynamics of the high-tech economy.

The stage in Teitelman's techno-industrial dramas is dominated by finance — its availability, cost and accompanying conditions. The role of money is repeatedly emphasized as the most influential variable in product development, manufacturing and marketing. "It is a mistake to view technologies as if they were isolated phenomena, rising as if by spontaneous combustion", he states. "Rather, they are built on a scientific base, shaped by the financial environment." And he goes on to argue that "Cheap, available capital can shrink the gap between science and technology, accelerate change, undermine the hegemony of the large corporations, and occasionally hold off the shadow of maturity. Large reserves of inexpensive financing can allow smaller, newer companies to exploit scientific pos-

sibilities earlier." If the science yields technology, as it so abundantly did in electronics, the small fry can prosper, Teitelman points out. But not so if the science remains recalcitrant, as has been the case so far in biotechnology — a scientific triumph yet to find great marketplace success.

As a case in point, Teitelman chronicles the postwar triumph of the transistor over the venerable vacuum tube, a victory produced by science-based technology, abundant venture capital and nimble management. Although "the vacuum tube industry was confident of its staying power" and even developed and marketed some transistor products, he says, "the smaller companies proved to be more aggressive marketers — they had no vacuum tube business to defend — and they were more effective in sniffing out and providing what the market needed and wanted".

In the postwar period, he observes, the unprecedented abundance of government research funds enriched the scientific and technological atmosphere, loosened a gusher of venture capital and spawned the creation of innumerable start-up companies that "surfaced in great waves, many with a technological mission and a zeal for competition. The ready availability of capital brought these new players more time to commercialize new technologies, to develop new products and skills (marketing, manufacturing, management) that had traditionally allowed larger corporations to preserve their position in a dangerous, unstable world."

Interacting with the availability of capital, he writes, is the state of the science that underlies the technology, and these factors, in turn, shape the corporate structure and managerial style of commercialization attempts. "Generally speaking, the ease with which a technology is commercialized varies with the maturity of its underlying science . . . The bureaucratic structure of most major drug companies and of drug regulation was formed, in part, as a response to the stubborn uncertainties of drug discovery and biology . . . The scarcity of venture capital in the early 1950s meant that television makers came predominantly from the ranks of prewar radio companies; the flood of equity capital in the 1980s floated many new computer and semiconductor firms, despite the daunting costs."

Teitelman is particularly incisive about the transition in research politics wrought by the postwar infusion of government money into academic science. From the late nineteenth century to the early 1950s, he points out, "the technological economy was dominated by a handful of giant companies: AT&T, General Electric, RCA, Westinghouse, Du Pont". Buttressed by economies of scale and ample corporate treasuries, these companies

"deployed their financial muscle to build patent fortresses, heavily influenced the government's R&D [research and development] and regulatory agendas, and, again through the power of the purse strings, held the greatest sway over academia". By the late 1950s, however, federal funds provided an alternative to industrial financing of university research, while small companies proved to be particularly fruitful settings for developing and commercializing electronic products.

As should be evident by now, Teitelman crams a great deal of fact, argument and interpretation into a modestly sized book. He is strong on insights, asking, for example, who on bedazzled Wall Street "really knew what was happening inside Texas Instruments not to mention at newer, smaller, more obscure firms?" The answer is that once share prices took off, few cared. But whether the grand themes of technology, finance and industry that he presents hold up under scrutiny is another matter.

Oddly enough, he concludes on a note of doubt, asserting: "The system that has developed in the United States is one of great complexity, constant change, and feedback. Quantification is difficult (though by now, great effort has been put into quantifying technological inputs and outputs), and one is left with description and portraiture."

True. Still, this is a commendable book, particularly worth reading in the current context of deliberate government efforts, in the United States and elsewhere, to cultivate high-tech riches. □

Daniel S. Greenberg is editor and publisher of Science & Government Report, 3736 Kanawha Street NW, Washington DC 20015, USA.

New in paperback

Irrationality: The Enemy Within by Stuart Sutherland. "A lively and readable book about modern follies" (Ernest Gellner, *Nature* **360**, 384; 1992). Penguin, £6.99.

Making Silent Stones Speak: Human Evolution and the Dawn of Technology by Kathy D. Schick and Nicholas Toth. Simon and Schuster, \$13. "Lively", "fascinating" and "up-to-date" (John Gower, *Nature* **362**, 672; 1993).

Genius: Richard Feynman and Modern Physics by James Gleick. Abacus, £8.99. A "moving, beautifully written, literate and perceptive account" (S. S. Schweber, *Nature* **360**, 375; 1992).

Science and Technology Advice to the President, Congress, and Judiciary (2nd edn) edited by W. T. Golden. AAAS Press, \$29.95. Contains 85 brief but informative essays, including one by McGeorge Bundy (see page 365 for his review of *James B. Conant*).

Theories of incompleteness

Thomas Banchoff

Holes and Other Superficialities. By Roberto Casati and Achille C. Varzi. MIT Press: 1994. Pp. 253. \$32.50, £26.90.

In case the potential reader has any doubt about the subject matter of *Holes*, the book jacket makes it clear — it is riddled with perforations. But what kind of holes are these? How do they differ from the holes in coffee cups, doughnuts and Swiss cheese? What characterizes their ‘holedness’? How do we recognize a hole? How do we describe a hole?

The authors are fascinated by these challenges. They develop a four-part philosophical theory of holes. How can we define what it means for something to have a hole (ontology)? How can we keep our words straight when we speak of the parts of a hole (mereology)? How can we distinguish and classify different types of holes (topology)? What does it mean to fill a hole (morphology)?

Already on page 7 the authors indicate the criteria against which their work should be evaluated: “A philosophical theory originates from astonishment and is judged on its ability to silence astonishment”. Although the book partly succeeds in heightening curiosity about the study of holes, it falls short of astonishing the nonspecialist, and insofar as it complicates descriptions and introduces many terms without adequate explanation, it fails to satisfy completely any curiosity it does stimulate. Many discussions are straightforward and easy to follow; others appear artificial and convoluted. Nonetheless, it raises several subtle questions that encourage the reader to think through his or her own examples, and it does provide the nonspecialist with an insight into the concerns and methodologies of philosophers.

A block of Emmental cheese is one of the author’s favourite “hosts” for holes left behind by bubbles that form during the cheesemaking process. A hole totally within the block is called a cavity, a hole that meets exactly one of the faces of the block in a single (circular) edge is called a hollow and a large bubble that meets two faces of the block in circular edges is called a tunnel. In a general object, the authors concentrate on hollows and tunnels delimited by edges that are discontinuities in the surface of the host, like the circular edges in the cheese. Moreover, in all the examples in the book, the edges of holes are planar objects.

Unfortunately, these restrictions limit the effectiveness of the theory, and the reader can often be distracted by thinking of examples that don’t fit.

Even for a block of Swiss cheese, it is not easy to see how the theory can deal with a bubble that cuts out a piece of an edge, or one that cuts off a corner. One of the key notions is that a hole is something that can be filled. We know how to fill a cavity, a tunnel or a hole in the cheese, but how do we fill these other holes? How do we fill the hole in a cup with a wavy lip or in a somewhat warped doughnut? How do we know where the hole begins? This is a crucial problem if, as the authors suggest, the essence of a hole is dependent on the way it meets the surface of its host (the superficiality referred to in the subtitle).

As a mathematician with a special interest in the geometry of lower and higher dimensions, I was initially pleased to see a reference to my favourite book, *Flatland* by Edwin A. Abbott (published in 1884, not 1882 as stated; and from which the pictures here are taken). It is a standard practice of geometers to develop concepts in ordinary space by first treating the analogous concepts in the plane. Unfortunately, the authors adopt a materialist attitude that prevents them from empathizing fully with the position of A Square, the planar protagonist of *Flatland*, investigating the analogue of a piece of Swiss cheese in his own dimension (represented, in some theories, by the top surface of the three-dimensional block of cheese, or some infinitesimally thin slice or cross-section of the block). A Square would argue that a circular bubble within the (square) slice is a cavity because it has

two boundary (curve) pieces, analogous to the distinguishing feature of a cavity in three-space. The authors, however, persist in looking at this hole from a three-space perspective and consider it to be a tunnel because they can make a loop of string in three-space that goes through the hole, as a loop would distinguish the tunnel in a doughnut. This misses the point of the dimensional analogy. One might just as easily get into an argument with a four-dimensional being who would consider a three-space cavity to be a tunnel from a four-space perspective.

The use of mathematical ideas of convexity and geometrical complements can go a long way towards avoiding philo-

sophical problems of ‘emptiness’. In defining holes purely in terms of what is missing, the authors consider holes to be “immaterial bodies”. To a mathematician, the situation is simpler. In any dimension, a convex object is defined by the property that any segment determined by two points of the object consists entirely of points of the object. Any object is contained in a smallest convex object, called its convex hull. Holes, then, are certain (maximally connected) pieces of the convex hull that are not parts of the original object. So a hole is a set of points just as the host is. There is nothing mysterious or ‘immaterial’ about the hole from this geometric viewpoint, although the reader will find many discussions of related philosophical problems in the book.

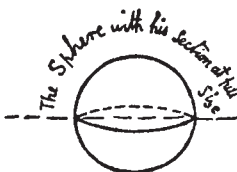
The diagrams in the book generally represent three-dimensional entities, so they are usually presented with some shading and perspective. Occasionally the authors use two-dimensional slices to illustrate their points, and this can lead to confusion in interpretation, as in the illustrations of a “perfect filling” of a hole and of a “canonical (non-privileged)” fillable part of a hole. There is no obvious way to visualize these as three-dimensional objects, and the concepts “canonical” and “privileged” are never properly defined.

Frequently the philosophical analysis causes the authors to deviate from common usage. For example, they state: “Holes can move, and regions of space cannot”. Why? An example will illustrate the process.

In a glass of water with a thick layer of oil on the top, as a bubble rises from the water to the interface between water and oil and finally into the oil, then in ordinary language one might say that it is the same bubble. On the other hand, in the language developed by the authors the bubble creates four different holes during its migration: first a cavity in the water; then two holes, a hollow in the water and another hollow in the oil; and finally a cavity in the oil.

The book ends with 66 or so theorems listed in an appendix along with various definitions and axioms. Despite this formalistic conclusion, the authors try to write for a general audience, and although they lapse into the jargon of one discipline or another at the ends of various chapters, it is always possible to pick up a fresh thread in the next. It is not, however, too easy to skip around: ideas are developed progressively and an innocent-sounding word may earlier have been given a precise, unexpected definition. □

Thomas Banchoff is in the Department of Mathematics, Brown University, Providence, Rhode Island 02912, USA.



Archetypes and ancestors

Colin Patterson

Richard Owen: Victorian Naturalist. By Nicolaas A. Rupke. Yale University Press: 1994. Pp. 462. £35.

To most contemporary biologists, Richard Owen (1804–92) is dimly recalled only as a villain, an eminent enemy of Darwin over whom T. H. Huxley won famous victories on Darwin's behalf. Comparative biologists should also know Owen as the originator of the distinction between homology and analogy, still the basis of systematics and the comparative method. And Owen has been in the news for naming Dinosauria, whose sesquicentennial was celebrated in 1991. There is ambiguity here between knave and hero, and ambiguity in Owen's nature and work is an underlying theme in Rupke's book.

There has been no biography of Owen since the conventionally eulogistic *Life* published by his grandson in 1894. For Rupke the reason is simple: "Owen was systematically written out of Victorian history by Darwin and his followers." Reinstatement has begun only in the past decade*, and has emphasized Owen's role in the context of evolution. Rupke, while reviewing the whole of Owen's career, stresses a different aspect, Owen's "drive for the creation of a national museum of natural history", in which he finds the key to understanding that career. After decades of politicking, Owen succeeded; the British Museum (Natural History) opened in 1881, with Owen (aged 76) its first director. To symbolize twentieth-century evaluation of Owen and the Darwinites, Rupke takes three statues in that museum, contrasting the blackened bronze of Owen's with the snowy marble of Darwin's and Huxley's. Yet there is irony here too. In today's museum (incognito as the "Natural History Museum", in the interests of marketing), visitors find Owen looking down the nave of his cathedral from the head of the stairs, a position he has occupied for 75 years. To get there, he displaced Darwin, unveiled in that position by the Prince of Wales in 1889 (Huxley joined Owen in the main hall in 1900). For most of my life, Darwin and Huxley flanked the main entrance and so were the first objects one encountered, but (in the interests of marketing) they are now tucked away beneath Owen's stairs, flanking a snack bar.

Rupke's book is not a conventional biography. For instance, there are only a

couple of casual references to Owen's wife (they were married for almost 40 years) and to his only child (who committed suicide in 1886, "jumping into the Thames, leaving his hat with purse, watch and address card inside it on the bank"). Instead, the book is a series of extended scholarly essays, each covering an aspect of Owen's professional career.

The first two essays are on "Museum Politics". Owen's ambiguity is emphasized here in his need to balance two forces, his patrons among the Oxbridge or



Enigmatic giant — Richard Owen in 1888.

Oxbridge-educated academic, aristocratic and clerical élite, and his urge to establish a power base in the metropolis, among his professional peers and superiors. Owen's university training was limited to six months at Edinburgh, which, as Adrian Desmond has shown, was the source of most radical theorizing in Victorian scientific London. Owen moved to London in 1825 and "worked his way up through a succession of apprenticeships", primarily in the Hunterian Museum of the Royal College of Surgeons, whence, already laden with honours, he moved to the British Museum in 1856. Owen had previously campaigned to get parts of the museum's collections incorporated in an enlarged Hunterian, as a national museum of comparative anatomy.

Having failed to get the museum to move to him, he moved to it and campaigned there for an independent national

museum of natural history. Pressure on space in the Bloomsbury building might be relieved by enlarging it, moving out the natural history collections or moving out the antiquities. The main reason for advocating the last option was the popularity of natural history with the majority of visitors — "the trade- and wage-classes" from the East End, close to Bloomsbury. But the power and tastes of the trustees and the principal librarian, Antonio Panizzi, were enough to maintain antiquities and the library on the central site and relegate natural history to "Albertopolis", the new museum complex that eventually arose in South Kensington, near the site of the Great Exhibition of 1851. Through all these negotiations, which occupied Owen for nearly four decades, Rupke argues that Owen's scientific research was guided by the needs of his museums, primarily for "icons", majestic or rare objects such as the giant extinct reptiles, birds and mammals, or *Archaeopteryx* (which Rupke consistently spells *-pterix*, evidently no slip, for the kiwi is also spelt *Apterix*). These icons attract the visitor, justify spacious premises, symbolize imperial prestige and encourage colonial natural history.

The central third of the book is occupied by two long essays on Owen's scientific work, divided by the themes of function, or Cuvierian method, and form, or the transcendental anatomy of Goethe, Oken and Geoffroy. Rupke argues for this division as reconciling Owen's political needs. With functionalism, teleology and the argument from design, he maintained the respect and patronage of Oxbridge clerisy and aristocracy; with transcendentalism he followed his own shifting allegiance towards the science of the continent and his Edinburgh-trained metropolitan colleagues, and towards transmutation. Here again, ambiguity is evident; Rupke calls it "the enduring epistemological duality of his oeuvre".

The fifth essay, "Eclipsed by Darwin", shows how, from the mid-1840s, Owen became an evolutionist (covertly supporting Robert Chambers, the anonymous "Vestigiarian", for example), and argues that the impulse came from zoogeography of living and extinct mammals. Yet the ambiguity persists, and although Owen

*Ospovat, D. *The Development of Darwin's Theory* (Cambridge University Press, 1981); Desmond, A. *Archetypes and Ancestors* (University of Chicago Press, 1984); Desmond, A. *The Politics of Evolution* (University of Chicago Press, 1989); Sloan, P. R. (ed.) *Richard Owen: The Hunterian Lectures in Comparative Anatomy, May–June 1837* (University of Chicago Press, 1992); Gruber, J. W. & Thackray, J. C. *Richard Owen Commemoration* (Natural History Museum Publications, 1992).

was infuriated by Darwin's counting him among creationists in the *Origin*, the error can be excused. Owen's wound was repaid in his anonymous review of the *Origin*, which opened an irreconcilable gulf between him and the Darwinites. The sixth and seventh essays cover the details of one stage in their revenge, the "hippocampus minor controversy" with Huxley over the relationship between man and apes. Huxley, who had been using Owen as a climbing frame for years, outclassed him as tactician and controversialist. The final essay covers minor topics (sea serpents, longevity, vivisection) used by Rupke to show how Owen sought to shift authority from the judiciary (in evidential standards), the clergy (in received wisdom) and the aristocracy (in inherited power) towards science, the real "ministry of truth", and how, once his ambition was achieved and he was installed in his cathedral in South Kensington, he was quite ready to argue against the patrons whom he had courted for so long.

Having spent the greater part of my life in the institution that Owen founded, and having read, used and admired substantial parts of his scientific work, I should be a ready convert to Rupke's "demythologising scholarship". My problem is Owen's prose. Rupke does not mention Owen's work on fishes, my own field. Here he is (in the 1860 textbook on palaeontology) summarizing the history of teleosts: "those species, such as the nutritious cod, the savoury herring, the rich-flavoured salmon, and the succulent turbot, have greatly predominated at the period immediately preceding and accompanying the advent of man; and that they have superseded species which, to judge by the bony Garpikes (*Lepisosteus*), were much less fitted to afford mankind a sapid and wholesome food". I have never found the prose or the ideas nutritious. Rupke also has trouble with Owen's style ("convoluted, evasive", "the fog of verbal obfuscation", "inimitable prose, resembling an obstacle course"). From the depth and range of his scholarship, Rupke has obviously spent years in Owen's company. Yet he has not come to like the man. At one point he paraphrases Hugh Trevor-Roper on Edmund Backhouse: "the mere fact that Owen praised a colleague did not necessarily imply self-glorification". Owen would (anonymously) review his own books and write that he had been "favoured to listen" to his own lectures. I finished the book feeling no closer to understanding Owen the man; he remains an enigmatic giant. He was a great morphologist, but he was also a politician, and as H. L. Mencken put it, a good politician is as unthinkable as an honest burglar. □

Colin Patterson is in the Department of Palaeontology, Natural History Museum, London SW7 5BD, UK.

Logic and laws

Ziauddin Sardar

The Rise of Early Modern Science: Islam, China and the West. By Toby E. Huff. Cambridge University Press: 1993. Pp. 409. £35, \$54.95.

Islamic Science and Engineering. By Donald Hill. Edinburgh University Press: 1994. Pp. 250. £39.95 (hbk); £16.95 (pbk).

QUESTIONS, the eleventh-century Muslim philosopher and scientist al-Baruni once said, have meaning only in the worldview and context within which they arise. The most often asked questions about the legacy of al-Baruni, Islamic science, is why it failed to produce a distinctively modern science. Similar questions have been raised about Chinese science. Joseph Needham, among others, has asked: "Given the superiority of Chinese science until the seventeenth century, why did China fail to produce a Galileo?"

Toby Huff sets out to answer these questions. From the eighth to the fourteenth centuries, he tells us, Islamic science was "the most advanced science in the world, greatly surpassing the West and China". Just what the achievements of Islamic science were is described by Donald Hill in his concise and masterly survey, which offers us a new synthesis based on recent research. The starting point for analysing the decline of Islamic science is the mathematical models of the fourteenth-century scientist ibn Shatir, and the work of astronomers at the famous observatory in Maragha, Adharbayjan, built in the thirteenth century by Nasir al-Din al-Tusi. According to Hill, the Maragha astronomers developed the Tusi couple and a theorem for the transformation of eccentric models into epicyclic ones. Copernicus not only used these two basic theorems to build his notion of heliocentricity but also used them at exactly the same points in the model. In other words, the lunar models of Copernicus and the Maragha school are identical. While Hill is concerned with discovering where Copernicus acquired the Maragha theory, Huff asks a much broader question. Why didn't ibn Shatir and his followers at Maragha take the leap to a heliocentric worldview? Why did the Muslims not go the last mile, especially when the Muslims had already taken on Ptolemy some two centuries earlier when ibn al-Haytham had boldly declared that "the arrangements proposed for planetary motions in Ptolemy's *Almagest* were 'false'"?

The answers, according to Huff, lie in the religious, sociological and cultural dimensions of the Muslim civilization. The blame rests squarely on Islamic law, Muslim family relationships, the universi-

ty system and theological disputes. We are told that because Islamic law "does not recognise corporate personalities" and "the idea of personal liability and the concept of negligence are unknown to Islamic law", cities and universities and other legally autonomous entities could not evolve in the Muslim world. Various schools of legal thought insisted on maintaining their separate identities and principles and therefore aborted the development of universal legal principles. The "extremely personalised nature of human relations" and the dominance of "the traditional extended family" meant that the Muslim civilization was not able to formulate "generalised universal norms". Moreover, "Islamic occasionalism" denied that the laws of nature were governed by a rational order. Also, science was essentially a marginal activity in Muslim civilization: the brilliance of its scientists was an exception to the rule rather than the norm and, in any case, society did not support the scientists, as many of them presented a threat to Islamic theory.

Frankly, Huff's diagnosis is patent nonsense. Islamic law may not recognize "personal negligence" in the sense of European law, but it is based on a much broader notion of individual responsibility and accountability. It does recognize corporate personalities: the *wuqfs*, the pious trusts and foundations that provided funding for universities, were a type of autonomous corporation. The banality of the suggestion that Islamic occasionalism or any other Islamic 'ism' denies the existence of a rational order defies comprehension. One definition of a Muslim is someone who reflects on and ponders the rational order of the Universe. This is why one-third of all the verses in the Koran are devoted to praising reason and exhorting the believers to use reason. Moreover, it defies all logic, not to say evidence put forward by such scholars as A. I. Sabra, whose arguments are dismissed by Huff with a sleight of hand, to suggest that science was a marginal activity in Islam. Marginal activities are not patronized by caliphs and sultans or supported by a network of magnificent libraries or have scientists and scholars criss-crossing the globe hunting out manuscripts and visiting research institutions. And above all, marginal activities do not take a people to the zenith of civilization.

To substantiate his intangible thesis, Huff systematically presents a totally distorted picture of Islam. We are, for example, told that in Islamic law murder is treated as "a private matter in which the state does not interfere". This will come as news to a billion Muslims across the world. The area in which the "central authority maintains a continuous vigilance", we are told further, is that of "crimes against god, the *hudud* crimes". As a secondary-school Muslim kid will tell

the Massachusetts sociologist, murder is a *hudd* crime. Islamic universities are dismissed as nonscientific institutions because, Huff alleges, contrary to much recent research, as charitable trusts they did not conduct research or hold examinations. But hospitals and observatories that were corporations, and according to Huff's own accounts, did conduct research and hold examinations, are given a primitive flavour and described as 'proto-scientific'. So nothing that is not Western can be described as scientific! Many of Huff's arguments about Islamic law are based on discredited works by orientalist — some, such as that of Goldziher, more than a hundred years old.

So what were the reasons for the decline of Chinese science? Here, Huff points at the alleged irrationality and complexity of the Chinese language which is not conducive to "clear and unambiguous communication" and is therefore not suitable for scientific inquiry. There are two problems with this suggestion. First, it was the same Chinese language that led to the development of Chinese science in the first place and to the discovery of the magnetic needle, the rudder, gunpowder and other technologies that came in rather handy for the evolution of modern science in Europe. Second, it was the same language that enabled modern China to acquire nuclear technology and seems to be no hindrance to Japan, which uses a parallel system and clearly leads the world in many fields of scientific inquiry.

From the irrationality of Chinese language we move to the implausibility of Chinese metaphysics. The Chinese notions of the organic world of primary forces (*yang* and *yin*), Huff tells us, is hardly a metaphysics worthy of the name. It does assume that "there is a pattern to existence in all things and that there is a unique way (*tao*) for all things, but the explanation of the patterns of existence is not to be sought in a set of laws or mechanical processes, but in the structure of the organic unit of the whole". The whole notion is little more than "a primitive but natural instinct of mankind" and, at best, it yields meaningless binary oppositions such as light and darkness. There is a total absence in Chinese thought of "a genuine dialectic of disputation and a faith in reason". Funny, then, how the Chinese managed to build a whole civilization on this 'primitive', 'natural' notion and evolved a science without any idea of reason that, according to Needham, "was much more efficient than the occidental in applying human natural knowledge to practical human needs" and in many ways "was much more congruent with modern science than was the world outlook of Christendom".

Both China and Islam have an integrated worldview. In their unique and individual ways, these cultures do not

place reason and ethics in two separate compartments. Although Islam extols the virtues of reason, it also warns against instrumental rationality — the intense debate in Muslim civilization between theologians and philosophers was a debate not about the use of reason but about the transformation of reason into an instrument of oppression. Both civilizations have a deep respect for nature: in Islam it is a sacred trust; in Chinese metaphysics it is an autonomous, integrated web; in both cases one approaches nature with respect and humility. Given the ecological devastation caused by modern science, both approaches have something going for them. And neither civilization accepts a totally mechanical view of the Universe. With hindsight, and a backhanded compliment to quantum physics, both can be said to be 'right'.



Mediaeval Arab astronomer by Dürer.

What Huff is really saying is that modern science did not develop in Islam and China because these civilizations failed to produce mercantile capitalism and large corporations, and refused to accept a deterministic model of the Universe and treat nature as an empire to be, as Bacon put it, "tortured" so that its secrets can be extracted. We are presented with a distinctively right-wing capitalist view of history that, while acknowledging the achievements of Islamic and Chinese science, tries to expunge them from the history of modern science. Huff's goal is to preserve the purity of Europe by demonstrating that the origins of modern science lie not in China and Islam but in "the fusion of Greek philosophy, Roman law" and that most rational of all metaphysical systems, "Christian theology".

While the rise of modern science is

presented by Huff as an exclusively European phenomenon, science itself is seen as bearing no cultural or social fingerprints of Europe. For a sociologist, Huff is dangerously unaware of recent advances in sociology of knowledge or the anthropology of science, and unfamiliar with the works of such philosophers of science as Paul Feyerabend, Imre Lakatos and Jerry Ravetz. Anything in Needham that does not fit his carefully constructed Eurocentric thesis is dismissed as inspired by Needham's Marxism. His notion of modern science as neutral, value-free, disinterested pursuit of uncontaminated truth is a pure fantasy.

Not surprisingly, the Europe of the Middle Ages that emerges from Huff's analysis is a heaven of rationality where religion had willingly taken a back seat, reason and conscience had been discovered for the first time and science advanced unhindered by dogma, persecution or other social and cultural impediments. In the twelfth century, we are told, Christianity in Europe had turned into a "corporation", a legal entity that did not really interfere with the work of the scientists — but, critics may ask, did it stop the emergence of the Inquisition a few centuries later? The "cultural outlook, social organization and economic performance" as well as "institutionalised disinterestedness and skepticism" of Europe prepared the ground for the emergence of modern science. The universities, far from being reluctant partners in scientific endeavour as is normally accepted, Huff contends, became the founding institutions of modern science. By the fifteenth century, the search for truth had become "part of the credo, the ethos, the cultural outlook of Europe". Of course, it was the same credo that launched Europe on its imperial adventure and colonization of the rest of the world. There is not even an iota of awareness in Huff's book that Europe is engaged in two conquests, that of nature and that of other cultures, at the same time. Nor is there an awareness that many postcolonial scholars, such as P. Petitjean and R. K. Kochhar, have demonstrated an intimate connection between the origins of modern science and the emergence of the European Empire. Universality of modern science is established as an empirical consequence of European expansion, not an epistemological cause of valid claims.

While *Islamic Science and Engineering* consolidates the recent advances in the history of Islamic science and technology, *The Rise of Early Modern Science* provides us with an excellent example of the emerging new strain of deep Eurocentrism. Huff is concerned not only with showing that modern science is its own origins, final end and self-perpetuating force, but also insists that non-Western cultures have no future unless students

from developing countries populate Western universities. (What they study in their own countries, in India, China and the Muslim world, tinged as it often is with their paranoid extended family systems and irrational metaphysics, can hardly be described as science!) Huff probably does not know that the Dogon people of West Africa, considered by anthropologists and orientologists to be one of the most primitive and irrational, knew many of the observations that Galileo's telescope made possible more than 1,500 years earlier. Two interesting questions arise. Did the Dogon invent some sort of telescope? Or did they have extraordinary eyesight? □

Ziauddin Sardar is a consulting editor of Futures, a monthly journal of forecasting, planning and policy.

Hypnosis revisited

H. J. Eysenck

From Mesmer to Freud: Magnetic Sleep and the Roots of Psychological Healing. By Adam Crabtree. Yale University Press: 1993. Pp. 413. \$50, £30.

CRABTREE is on the faculty of the Center for Training in Psychotherapy in Toronto, Canada, and a psychotherapist in a private practice. He is concerned to trace the historical roots of the practice of psychotherapy from Mesmer to Freud, and his historical knowledge is considerable; he cannot be faulted on the factual account he gives of Mesmer and his disciples, critics and followers. I had to look at the same literature when I was experimenting with hypnosis and suggestibility, and found it both fascinating and amusing. Through it all runs the Cartesian dilemma: if body and mind are separate entities, how can the mind influence the body? How can purely verbal manipulation of the mind (suggestion), as for instance in hypnosis, cause warts to disappear, breasts to grow larger, major surgery to be carried out without any feeling of pain? The answer of course is that Descartes was wrong; just as physicists had to learn to deal with a space-time continuum, so we will have to learn to deal with a body-mind continuum. But it took some time to appreciate this, and the history of how the role of suggestion was camouflaged as the effect of 'magnetic fluids' is long.

The famous Benjamin Franklin Commission in Paris, which was convened to judge Mesmer's claims on a scientific basis, hit the nail on the head. There were clear-cut cases of apparently miraculous cures of bodily ailments, but magnetic fluids had nothing to do with them; 'imagination' (suggestibility) was the important ingredient. The patients who were

cured of a wide variety of often severely incapacitating diseases were very probably suffering from hysterical symptoms rather than physical ailments; these gross symptoms (hysterical blindness, lameness and so on) were fairly common before the First World War; they seem to have died out, and are hardly ever seen today. Of course, there were never any control groups, and there was little follow-up; neither was there any attention to failures, which were probably numerous. We cannot take the case histories as true accounts, either; quite likely they have grown in the telling. Even Freud's later case histories are known to have been economical with the truth; his fabled 'cures', as in the case of the Wolf Man, were not cures at all. The Wolf Man, after Freud's 'cure', continued for 60 years to show the same symptoms as before, despite much further therapy.

Crabtree tells the story well as it unrolls through names famous in the history of hypnosis, such as Puységur, Charcot, Binet, Braid and many others. They all made their contributions and discovered the main phenomena in this field. The Franklin Commission had already noted that a close bond seemed to develop between the magnetizer and the magnetized — shades of Freud's notions of 'transference'. Puységur used to interpret the dreams of his patients, and found their own interpretation useful. Many critics noted the tendency of young women (who probably constituted the majority of patients) to develop sexual affinities with their therapists, and such spiritual affinities do not always seem to have remained spiritual. Here again there are close parallels nearer home; many psychotherapists today seem to regard their profession as giving them a kind of *droit de seigneur* over attractive female patients; there are many cases coming up before ethical committees, or even judges. *Plus ça change* . . . Finally, Mesmer brooked no disagreement among his followers and, like Freud, forced the leading spirits among them to split off and found their own movements. Truly history repeating itself, the second time as farce.

In a book such as this one might have expected some discussion of our modern views about the astonishing phenomena encountered. Nothing of the kind. The historical account trickles on, but no explanation is attempted of the 'cures' reported or of the many other details so painstakingly uncovered. We may presume that most if not all of the phenomena involved concern placebo effects, but oddly enough the word does not appear in the index. This is a pity: placebo effects are so universal and so strong that a thorough discussion would have been invaluable. An example is a recent study of medical treatments of several serious physical conditions, all of which had been in use over

several years but had then been abandoned because proof had been forthcoming that they had no physical effect. The study showed that in 70 per cent of all cases where these useless treatments had been used, the treatments had in fact had excellent or very good effects. Another recent report compared Valium with a placebo in the treatment of anxiety and found no difference. When we consider that medical treatment until the time of Pasteur probably did more harm than good, we may conclude that the prestige of medicine is largely built on placebo effects — and in many cases probably continues along similar lines.

If this is true of physical diseases, it seems likely to be even more true in psychiatric disorders. When I pointed out in 1952 that psychoanalytic treatment had not been shown to have better outcomes in cases of severe neurotic disorders than no treatment or placebo treatment, orthodoxy was up in arms. Yet only recently a meta-analysis of the latest 19 studies comparing psychoanalytic treatment with no treatment at all reported that in none of the studies was there any evidence that the neurotics treated with psychoanalysis were any better a year after treatment was finished than the controls who had no treatment at all. The whole history of psychotherapy from Mesmer to Freud is a long list of theoretical inventions, from magnetic fluids to Oedipus complexes, to usurp the position that really should go to the placebos. In another large-scale meta-analysis, none of the most widely used psychotherapies did significantly better than placebo treatments.

Perhaps Crabtree as a psychotherapist did not feel inclined to discuss such matters; psychotherapists are not usually too happy with such facts. But by not even mentioning placebo treatment, Crabtree leaves out the Prince from Hamlet's tragedy; he makes it impossible for himself to offer any kind of explanation for the odd and extraordinary events he recounts so well. It is difficult to guess what readers not versed in these matters will make of the account; when they find that spiritualism, table-turning and psychic research (of the pre-scientific era) are drawn into the cauldron of speculation, they may well feel that we have left science behind and that Franklin and his commission really should have the last word — 'imagination' is at the back of it all. The pity is that only recently has there been much interest in the scientific study of suggestibility (although Binet made a brave start at the beginning of this century) and that Crabtree has nothing to say on the subject. Two cheers for a good history, but none for a lack of science. □

H. J. Eysenck is in the Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.